

2m11.3223.17

Université de Montréal

Analyse structurale de la sélectivité de la liaison à l'ADN par les récepteurs des oestrogènes  
et des glucocorticoïdes

par  
Geneviève Pesant

Département de Biochimie  
Faculté de Médecine

Mémoire présenté à la Faculté des études supérieures  
en vue de l'obtention du grade de Maître ès sciences  
en biochimie

avril, 2004

© Geneviève Pesant, 2004



W

4

U58

2005

v. 015

**Direction des bibliothèques**

**AVIS**

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

**NOTICE**

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal  
Faculté des études supérieures

Ce mémoire intitulé :

Analyse structurale de la sélectivité de la liaison à l'ADN par les récepteurs des oestrogènes  
et des glucocorticoïdes

présenté par :  
Geneviève Pesant

a été évalué par un jury composé des personnes suivantes :

Président-rapporteur : Dr Pascal Chartrand  
Directeur de recherche : Dr Serguei Chteinberg  
Codirectrice : Dr Sylvie Mader  
Membre du jury : Dr Normand Brisson

## RÉSUMÉ

Cette étude avait pour but de déterminer le rôle des acides aminés de la boîte P du récepteur des oestrogènes (ER, NR3A1) et du récepteur des glucocorticoïdes (GR, NR3C1) dans le positionnement de l'hélice de reconnaissance de l'ADN (H1) dans le sillon majeur de l'ADN. Pour ce faire, une analyse des structures tridimensionnelles des complexes récepteur-ADN ainsi que des expériences de modélisation moléculaire ont été effectuées. Il a été démontré que le positionnement correct de H1 dans le sillon majeur de l'ADN était un facteur déterminant de la sélectivité de la liaison à l'ADN. Il a aussi été démontré que l'empaquetage de Gly26 ou de Ser26 sur deux arginines dictait le positionnement de la partie N-terminale de H1. Un modèle intégrant le rôle des acides aminés impliqués dans la sélectivité de la liaison à l'ADN et le positionnement de H1 dans le sillon majeur de l'ADN est proposé.

**Mots clés :** interaction protéine-ADN, boîte P, modélisation moléculaire, hélice alpha, récepteur nucléaire.

## ABSTRACT

The goal of this study was to determine the role of the P box amino acids from the oestrogen receptor (ER, NR3A1) and from the glucocorticoid receptor (GR, NR3C1) in the positioning of the DNA recognition helix (H1) in the DNA major groove. To achieve this, the three-dimensional structures of the receptor-DNA complexes were analyzed and molecular modeling experiments were conducted. It has been demonstrated that the correct positioning of H1 in the DNA major groove was a determining factor in the selectivity of DNA binding. It has also been demonstrated that the packing of Gly26 or Ser26 on two arginines dictated the positioning of H1 N-terminus. A model integrating the role of amino acids implicated in the DNA binding selectivity and the positioning of H1 in the DNA major groove is proposed.

**Keywords :** protein-DNA interaction, P box, molecular modeling, alpha helix, nuclear receptor.

## TABLE DES MATIÈRES

<b>RÉSUMÉ</b> .....	iii
<b>ABSTRACT</b> .....	iv
<b>TABLE DES MATIÈRES</b> .....	v
<b>LISTE DES TABLEAUX</b> .....	ix
<b>LISTE DES FIGURES</b> .....	x
<b>LISTE DES ABRÉVIATIONS</b> .....	xii
<b>LISTE DES CODES DES ACIDES AMINÉS</b> .....	xiii
<b>DÉDICACE</b> .....	xiv
<b>REMERCIEMENTS</b> .....	xv
<b>CHAPITRE 1 : INTRODUCTION</b> .....	1
<b>1.1 Les ligands associés aux récepteurs nucléaires</b> .....	2
<b>1.1.1 Structure et effets médiés par les ligands des récepteurs nucléaires</b> .....	2
<b>1.1.2 Mécanismes d'action des ligands et de leur récepteur</b> .....	4
<b>1.1.3 Spécificité d'action tissulaire et génique</b> .....	5
<b>1.2 La superfamille des récepteurs nucléaires</b> .....	5
<b>1.2.1 Domaines fonctionnels des récepteurs nucléaires</b> .....	7
<b>1.2.2 Les éléments de réponses aux hormones</b> .....	8
<b>1.2.3 Modes de dimérisation</b> .....	9
<b>1.3 Domaine de liaison à l'ADN</b> .....	11
<b>1.3.1 Structures secondaire et tertiaire du DBD</b> .....	11
<b>1.3.2 Interaction avec l'ADN</b> .....	13
<b>1.4 Sélectivité de la liaison à l'ADN</b> .....	15
<b>1.4.1 La boîte P vs les paires de bases centrales des HREs</b> .....	17
<b>1.4.2 Autres facteurs jouant un rôle dans la sélectivité de la liaison à l'ADN</b> ...	18

1.5	Code de reconnaissance de l'ADN.....	18
1.6	Questions abordées dans ce mémoire.....	19
<b>CHAPITRE 2 : RÉSULTATS, 1<sup>ÈRE</sup> PARTIE (ARTICLE).....</b>		<b>22</b>
2.1	Contribution des auteurs.....	24
2.2	Summary.....	25
2.3	Introduction.....	25
2.4	Materials and methods.....	28
2.4.1	Plasmids.....	28
2.4.2	E. coli expression of the ER $\alpha$ and ER $\beta$ DNA binding domain and derivatives thereof.....	29
2.4.3	Immunoblot analyses.....	29
2.4.4	Gel shift assays.....	30
2.4.5	CAT assays.....	31
2.4.6	Modeling.....	31
2.5	Results.....	32
2.5.1	The patterns of binding of both ERs to elements containing base replacements in both motifs are similar to each other and to those of the isolated DBDs.....	32
2.5.2	Point mutations in the consensus ERE destabilize interaction with ERs according to chemical compatibility and steric constraints with interacting amino acids.....	34
2.5.3	Roles of Glu 203 and Lys 210 in the specificity and affinity of ER interaction with palindromic response elements.....	38
2.5.4	The effects of mutations of Glu 203 on selective DNA binding are not predictable directly from chemical and steric compatibility rules.....	42



2.5.5	Lys 210 maintains binding to the consensus EREs with mutants at position 203 and prevents binding of mutant E203R to novel response elements.....	47
2.6	Discussion.....	48
2.7	Acknowledgements.....	52
2.8	References.....	52
<b>CHAPITRE 3: MATÉRIEL ET MÉTHODES; RÉSULTATS, 2<sup>ÈME</sup> PARTIE.....</b>		<b>54</b>
3.1	Matériel et méthodes.....	55
3.2	Résultats.....	56
3.2.1	Différences dans le positionnement de l'hélice de reconnaissance de l'ADN (H1) du GRDBD et du ERDBD dans le sillon majeur de l'ADN.....	56
3.2.2	Conséquences d'un changement de position de H1 pour la liaison à l'ADN.....	56
3.2.3	Acides aminés responsables du positionnement de H1 dans le sillon majeur de l'ADN.....	59
<b>CHAPITRE 4 : DISCUSSION.....</b>		<b>65</b>
4.1	Modèle proposé expliquant le mécanisme de la sélectivité de la liaison à l'ADN par les récepteurs stéroïdiens.....	66
4.1.1	Le modèle de Suzuki.....	66
4.1.2	Modèle proposé tenant compte du positionnement de H1 dans le sillon majeur de l'ADN.....	66

4.1.2.1	Rôle des acides aminés en deuxième position de la boîte P.....	69
4.1.2.2	Rôle des acides aminés en troisième position de la boîte P.....	70
4.1.2.3	Rôle des acides aminés en première position de la boîte P.....	71
4.2	Limites du modèle proposé.....	72
4.3	Limites de la méthode.....	74
4.4	Conclusion et perspectives.....	74
<b>BIBLIOGRAPHIE.....</b>		<b>76</b>

## LISTE DES TABLEAUX

### CHAPITRE 1 :

**Tableau I :** Liste de quelques récepteurs nucléaires de mammifères.....6

**Tableau II :** Exemples de séquences naturelles liées par le ER.....10

### CHAPITRE 3 :

**Tableau III :** Ponts H existants entre le ERDBD et le ERE en fonction de la position  
de H1 du ERDBD.....59

## LISTE DES FIGURES

### CHAPITRE 1 :

<b>Figure 1.1 :</b>	Structure de quelques ligands connus.....	3
<b>Figure 1.2 :</b>	Mécanismes d'action des ligands et des récepteurs nucléaires.....	4
<b>Figure 1.3 :</b>	Domaines fonctionnels d'un récepteur nucléaire typique.....	7
<b>Figure 1.4 :</b>	Configurations possibles des demi-sites liés par les récepteurs nucléaires....	9
<b>Figure 1.5 :</b>	Séquence et structure secondaire du GRDBD et du ERDBD.....	12
<b>Figure 1.6 :</b>	Structure tertiaire du DBD des récepteurs stéroïdiens.....	13
<b>Figure 1.7 :</b>	Structure du complexe DBD-HRE et détails de l'interaction entre les DBDs du ER et du GR avec l'ADN.....	16
<b>Figure 1.8 :</b>	Schéma des contacts directs entre les acides aminés du GRDBD et du ERDBD et leur HRE respectif.....	17

### CHAPITRE 2 :

<b>Figure 2.1 :</b>	Effect of replacements in both arms of the ERE palindrome on complex formation with full length ERs and isolated DBDs.....	33
<b>Figure 2.2 :</b>	Model for the selectivity of steroid receptor/DNA interactions.....	35
<b>Figure 2.3 :</b>	Chemical and steric constraints determine the specificity of complex formation with wt ER $\alpha$ .....	36
<b>Figure 2.4 :</b>	ERs bind with high selectivity to PuGNNCA palindromes.....	39
<b>Figure 2.5 :</b>	Role of K210 and E203 in the affinity and specificity of receptor-DNA complex formation.....	41
<b>Figure 2.6 :</b>	Mutations of Glu 203 do not generate the expected switch in DNA binding specificity.....	43
<b>Figure 2.7 :</b>	Mutation Lys 210 Ala introduced in Glu 203 mutants reduces binding to the consensus ERE and reveals the switched specificity of mutant E203R.....	44
<b>Figure 2.8 :</b>	Altered patterns of reporter vector transactivation by ER mutants.....	45
<b>Figure 2.9 :</b>	Modeling of the effect of mutations at position 203 on interaction with palindromic response elements.....	46

### CHAPITRE 3 :

- Figure 3.1 :** Différences dans la position de H1 du GRDBD et du ERDBD alors que les paires de bases centrales de leur HRE respectif sont superposées.....57
- Figure 3.2 :** Collisions entre les atomes des acides aminés du GRDBD et les atomes des bases du GRE causées par le déplacement de H1 du GRDBD dans la position occupée par H1 du ERDBD.....58
- Figure 3.3 :** Entassement de Ser26 sur Arg56 et Arg63 dans le GRDBD.....60
- Figure 3.4 :** Entassement de Gly26 sur Arg56 et Arg63 dans le ERDBD.....61
- Figure 3.5 :** Différences de positionnement des H1 du GRDBD et du ERDBD lorsque Arg56, Arg63 et le groupement phosphate de G-5 sont superposés.....62
- Figure 3.6 :** Déplacement de H1 du ERDBD après le remplacement de Arg56 et Arg63 par des alanines.....63
- Figure 3.7 :** Déplacement d'une partie de H1 du ERDBD après le remplacement de Ala29 par une glycine.....64

## LISTE DES ABRÉVIATIONS

A	Adénine
ADN	Acide désoxyribonucléique
AF	Fonction d'activation de la transcription
AR	Récepteur des androgènes
C	Cytosine
DBD	Domaine de liaison à l'ADN
DR	Répétition directe
ER	Récepteur des oestrogènes
ERE	Élément de réponse aux oestrogènes
ERR	Récepteur relié au récepteur des oestrogènes
G	Guanine
GR	Récepteur des glucocorticoïdes
GRE	Élément de réponse aux glucocorticoïdes
H1	Hélice de reconnaissance de l'ADN
HRE	Élément de réponse aux hormones
Hsp	Protéine de choc thermique
IP	Palindrome inversé
LBD	Domaine de liaison au ligand
MR	Récepteur des minéralocorticoïdes
Pont H	Pont hydrogène
PR	Récepteur des progestines
Pu	Purine (A ou G)
RAR	Récepteur de l'acide rétinoïque
RMSD	Root mean square deviation
SF-1	Facteur stéroïdogénique 1
T	Thymine
TR	Récepteur de l'hormone thyroïdienne
VDR	Récepteur de la vitamine D

**LISTE DES CODES DES ACIDES AMINÉS**

A	Ala	Alanine
C	Cys	Cystéine
D	Asp	Acide aspartique
E	Glu	Acide glutamique
F	Phe	Phénylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Méthionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Sérine
T	Thr	Thréonine
V	Val	Valine
W	Trp	Tryptophane
Y	Tyr	Tyrosine

## DÉDICACE

*Ce mémoire est dédié à ma mère, Lorraine Pesant, née Soucy, 1953-1999, qui a quitté ce monde tout juste avant le début de mes études universitaires. Son rêve était que ses enfants puissent faire les études de leur choix. Ce mémoire représente donc l'accomplissement de ce rêve. Merci Lorraine de m'avoir donné la possibilité de faire des études universitaires et de m'avoir accompagnée dans mes pensées tout au long de ces cinq années.*

« I would rather have had one breath of her hair, one kiss from her mouth, one touch of her hand, than eternity without it. One. »

- *City of Angels*, 1998



## REMERCIEMENTS

*Je voudrais remercier ma famille et mes amis pour leur support continu pendant mes études universitaires ainsi que mes collègues biochimistes Sanny Moussette, Jean-Nicolas Mess, Christian Jossart, Julie Simard, Karine Boulay et Frédérick-Antoine Mallette pour avoir partagé avec moi les hauts et les bas des études universitaires.*

*Un merci très spécial à Marie Pageau pour son aide, ses trucs et conseils ainsi que pour son support moral et son écoute qui ont été plus que nécessaires.*

*Je voudrais aussi remercier mes collègues du 'RNA lab', en particulier César Gomez-Aguilera, Ekaterina Zagriadskaïa, Sandra Chaker, Matthieu Gagnon et Félix Doyon, pour leur aide et leurs conseils pour la réalisation de mon projet de recherche.*

*Je tiens à remercier sincèrement ma co-directrice, Dr Sylvie Mader, pour m'avoir introduite dans le domaine fascinant du récepteur des oestrogènes et des récepteurs nucléaires, m'avoir permis de collaborer avec son équipe de recherche et pour son aide précieuse lors de la rédaction de ce mémoire.*

*Je voudrais aussi remercier le personnel de soutien du Département de Biochimie de l'Université de Montréal, en particulier Sylvie Beauchemin, pour leur aide technique et leur excellent travail malheureusement pas toujours reconnu à sa juste valeur.*

*Enfin, merci à mon directeur de recherche, Dr Serguei Chteinberg, pour m'avoir permis de réaliser mon projet de recherche au sein de son laboratoire.*

## **CHAPITRE 1 : INTRODUCTION**

Les effets des hormones stéroïdiennes sur le corps humain sont relativement bien connus et peuvent affecter la vie de tous les jours. Il suffit de penser aux menstruations, au syndrome prémenstruel, à la puberté, à la grossesse ou à la pilule contraceptive. Les hormones stéroïdiennes sont aussi un sujet d'actualité lorsqu'on évoque le dopage dans les sports et les traitements aux hormones (ou dérivés) utilisés pour soigner une maladie génétique, le syndrome d'insensibilité aux androgènes par exemple, ou pour traiter les symptômes de la ménopause. Dans ce dernier cas, les études s'accumulent mais se contredisent sur l'impact des oestrogènes utilisés dans ces traitements sur l'incidence du cancer du sein. Ceci démontre à quel point nos connaissances dans ce domaine sont insuffisantes. Il faut donc continuer les recherches pour mieux comprendre les mécanismes de fonctionnement des hormones stéroïdiennes et de leurs récepteurs spécifiques. Ce mémoire, qui a pour sujet plus spécifiquement les récepteurs des glucocorticoïdes et des oestrogènes, se veut une contribution à l'avancement des connaissances dans le domaine des hormones et récepteurs stéroïdiens.

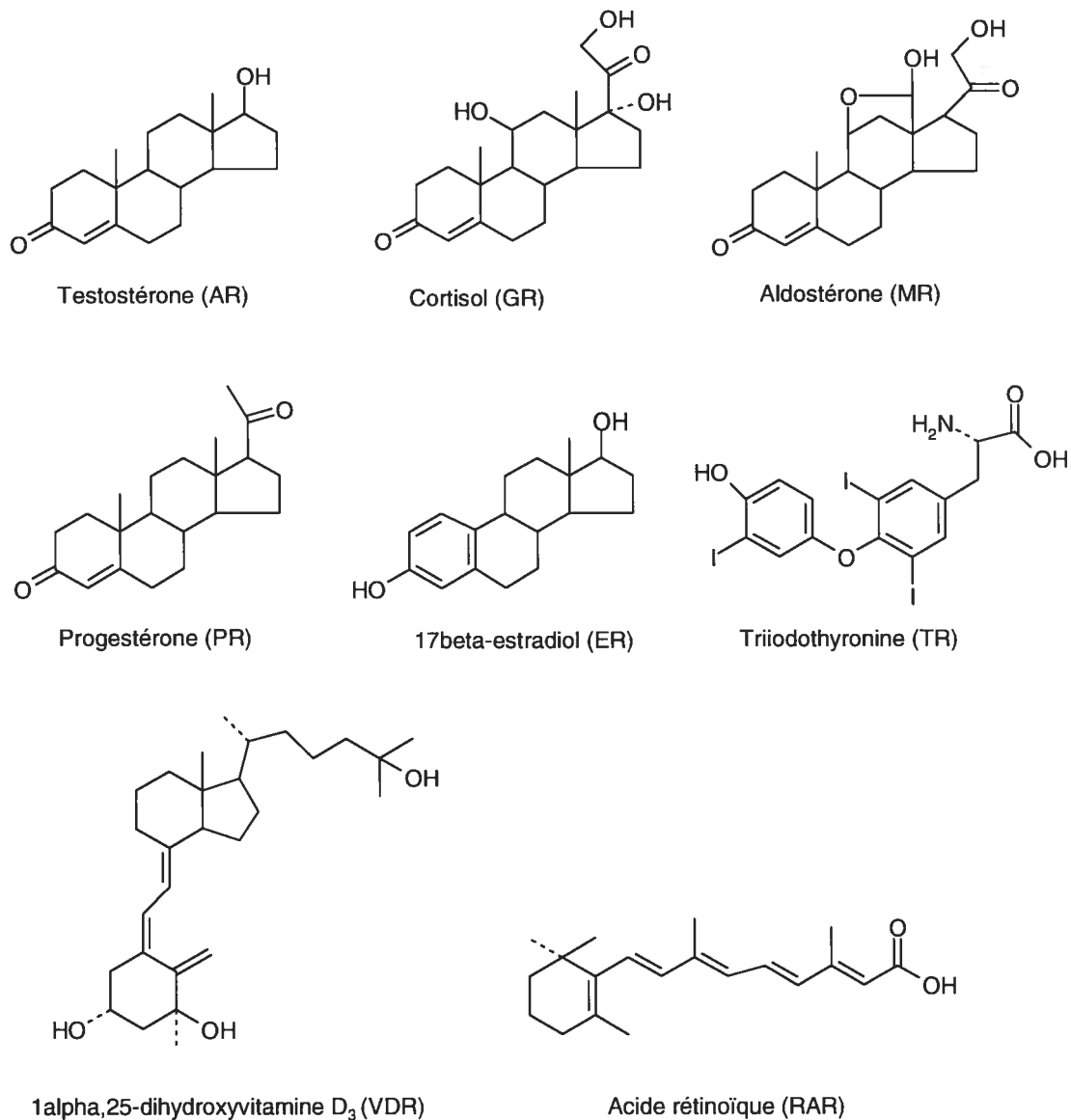
## **1.1 Les ligands associés aux récepteurs nucléaires**

Les hormones stéroïdiennes, qui sont les ligands les plus connus des récepteurs nucléaires, forment en fait un sous-groupe de molécules faisant partie d'un ensemble beaucoup plus large de ligands lipophiles. Ces ligands sont regroupés ensemble surtout parce que les récepteurs de ces ligands forment la superfamille des récepteurs nucléaires, qui sera décrite plus en détails à la section 1.2.

### **1.1.1 Structure et effets médiés par les ligands des récepteurs nucléaires**

Ces ligands sont généralement de petites molécules lipophiles ce qui leur permet de traverser la membrane cellulaire pour interagir avec leur récepteur spécifique [1]. La structure des plus connus de ces ligands est présentée à la figure 1.1. Ce groupe de ligands contient les hormones stéroïdiennes (ex. : glucocorticoïdes [2]), les hormones thyroïdiennes, les rétinoïdes et la vitamine D entre autres. Les processus physiologiques dans lesquels sont impliqués ces ligands sont variés : développement, croissance, reproduction, différenciation cellulaire, prolifération cellulaire, apoptose et maintien de

l'homéostasie. Des dérivés de ces ligands sont utilisés pour la contraception, le contrôle de l'inflammation, la prévention de l'ostéoporose et pour le traitement de plusieurs maladies telles que le diabète, les maladies de la peau, les syndromes de résistance aux hormones et certains cancers [1, 3]. La diversité et la complexité des processus dans lesquels est impliquée cette famille de ligands nous rappellent encore une fois l'importance de bien comprendre les mécanismes par lesquels ces ligands transmettent les messages d'une cellule à l'autre ou d'un tissu à l'autre.



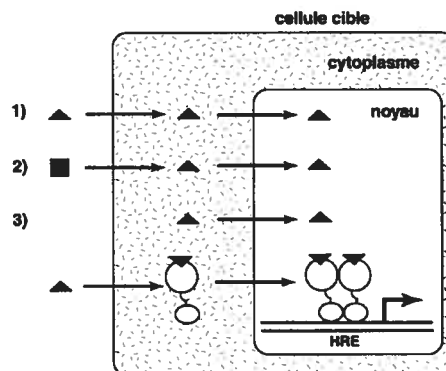
**Figure 1.1 : Structure de quelques ligands connus.** Le récepteur nucléaire spécifique à chacun des ligands est indiqué entre parenthèses.

### 1.1.2 Mécanismes d'action des ligands et de leur récepteur

Récemment, il a été démontré que l'estradiol pouvait avoir des actions non génomiques sur la cellule, c'est-à-dire des effets cellulaires ne découlant pas directement d'un changement dans l'expression des gènes [4]. De plus, les récepteurs nucléaires peuvent agir de façon indépendante du ligand [5]. Ces deux mécanismes d'action, quoique importants, n'entrent pas dans le cadre de ce mémoire. Nous ne nous attarderons donc qu'à décrire les mécanismes dépendants du ligand dont les effets sont génomiques.

La figure 1.2 décrit les principaux mécanismes de génération des ligands des récepteurs nucléaires : 1) le ligand provient d'un organe du système endocrinien où il a été synthétisé ou de l'environnement et entre dans la cellule cible après avoir été amené par la circulation sanguine, 2) le ligand peut être synthétisé dans la cellule cible à partir d'un précurseur et 3) le ligand peut être un métabolite synthétisé *de novo* dans la cellule cible [5].

Les récepteurs « nucléaires » peuvent être localisés soit dans le cytoplasme ou dans le noyau en absence de ligand. Dans certains cas, le récepteur est en association avec un large complexe protéique contenant des protéines de choc thermique (par exemple Hsp90 ou Hsp56), des chaperonnes, etc. Ce complexe protéique maintient le récepteur sous forme monomérique, empêchant la liaison à l'ADN. La liaison du ligand au récepteur entraîne un changement de conformation du récepteur, qui le libère du complexe protéique permettant ainsi la liaison du récepteur à l'ADN. Dans d'autres cas, le récepteur est déjà en association avec l'ADN dans un état transcriptionnellement inactif, et devient actif suite à l'association avec le ligand [4, 5].



**Figure 1.2 : Mécanismes d'action des ligands et des récepteurs nucléaires.** Les chiffres 1, 2 et 3 représentent les différentes sources de ligand. Se référer au texte pour les détails.

Les récepteurs nucléaires, une fois actifs, lient l'ADN sur des éléments de réponse aux hormones (HREs). Ces éléments de réponses seront décrits en détails à la section 1.2.2. Lorsque liés à l'ADN au niveau du promoteur de gènes cibles, les récepteurs nucléaires peuvent soit induire ou réprimer l'expression du ou des gènes cibles en recrutant différents cofacteurs (coactivateurs ou corépresseurs). Les coactivateurs peuvent modifier l'état de la chromatine et faciliter le recrutement des facteurs de transcription de base [4, 5].

### **1.1.3 Spécificité d'action tissulaire et génique**

La complexité et les différentes variantes du mécanisme d'action des ligands des récepteurs nucléaires font en sorte que la spécificité d'action des ligands est assurée. Les différents facteurs impliqués dans la spécificité de la réponse cellulaire ou tissulaire aux ligands sont [6]:

- les variations histo-spécifiques dans le métabolisme des ligands,
- la distribution tissulaire des différents récepteurs nucléaires,
- la distribution tissulaire des différents coactivateurs et corépresseurs,
- la coopérativité entre les récepteurs nucléaires et les autres facteurs de transcription fixés sur les différents promoteurs et
- l'état de la chromatine pour chaque élément de réponse.

## **1.2 La superfamille des récepteurs nucléaires**

Les récepteurs nucléaires sont, comme nous l'avons vu à la section 1.1.2, des facteurs de transcription ligand dépendants soupçonnés de dériver d'un ancêtre commun [7, 8]. Des études évolutives ont permis la séparation de la superfamille des récepteurs nucléaires en six classes qui sont présentées au tableau I [5]. Ce tableau démontre la grande diversité des membres de la famille des récepteurs nucléaires. Plusieurs n'ont pas encore de ligand connu, et sont appelés récepteurs orphelins. Dans cette section, nous ne nous attarderons pas sur chacun des récepteurs nucléaires mais nous décrirons plutôt les caractéristiques communes à cette famille de récepteurs.

**Tableau I : Liste de quelques récepteurs nucléaires de mammifères.** Les récepteurs nucléaires sont regroupés selon leurs classes respectives, indiquées par un chiffre romain. Un minimum d'un récepteur pour chaque classe est décrit. Pal, palindrome; DR-*n*, répétition directe; IP-*n*, palindrome inversé; *n*, nombre de paire de bases entre les deux demi-sites. Ht, hétérodimère; Mo, monomère; Hm, homodimère.

Classe	Récepteur	Sous-type	Nom du récepteur	Ligand	Élément de réponse	Dimérisation
I	TR	$\alpha, \beta$	Récepteur de l'hormone thyroïdienne	Hormone thyroïdienne	Pal, DR-4, IP	IIt
	RAR	$\alpha, \beta, \gamma$	Récepteur de l'acide rétinolique	Acide rétinolique	DR-2, DR-5, Pal, IP	IIt
	VDR		Récepteur de la vitamine D	1-25(OH) <sub>2</sub> vitamine D <sub>3</sub>	DR-3, IP-9	IIt
	CAR	$\alpha, \beta$	Récepteur constitutif des androstanes	Androstanes	DR-5	IIt
	FXR		Récepteur des farnésolides X	Acides biliaries	DR-4	IIt
	ROR	$\alpha, \beta, \gamma$	Récepteur orphelin relié au RAR	Inconnu	Demi-site	Mo
II	UR		Récepteur ubiquitaire	Inconnu	DR-4	IIt
	RXR	$\alpha, \beta, \gamma$	Récepteur de l'acide rétinolique X	Acide 9- <i>cis</i> -rétinolique	Pal, DR-1	IIm
	PNR		Récepteur nucléaire spécifique au photorécepteur	Inconnu	DR-1, demi-site	Mo, IIm
	GR		Récepteur des glucocorticoïdes	Glucocorticoïdes	Pal	IIm
	AR		Récepteur des androgènes	Androgènes	Pal	IIm
	MR		Récepteur des minéralocorticoïdes	Minéralocorticoïdes	Pal	IIm
III	ER	$\alpha, \beta$	Récepteur des oestrogènes	Estradiol	Pal	IIm
	PR		Récepteur des progestines	Progestines	Pal	IIm
	ERR	$\alpha, \beta, \gamma$	Récepteur relié au ER	Inconnu	Pal, demi-site	Mo, IIm
IV	NGF1-B	$\alpha, \beta, \gamma$	Clone B induit par les facteurs de croissance des nerfs	Inconnu	Pal, DR-5	Mo, IIm, IIt
V	SF-1	$\alpha, \beta$	Facteur stéroïdogénique 1	Oxystérols	Demi-site	Mo
VI	GCNF		Facteur nucléaire des cellules germinales	Inconnu	DR-0	IIm
0	SIIP		Petit partenaire hétérodimérique	Inconnu	-	IIt

### 1.2.1 Domaines fonctionnels des récepteurs nucléaires

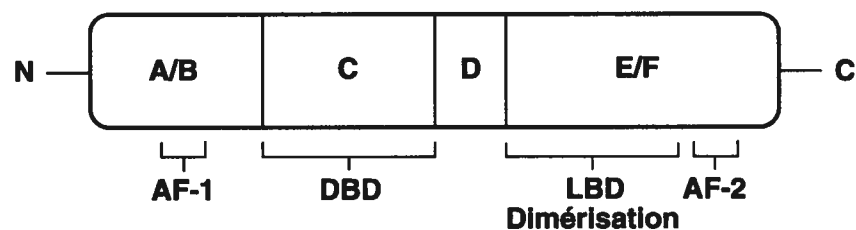
Les récepteurs nucléaires possèdent une structure modulaire et leurs domaines fonctionnels sont indépendants l'un de l'autre. Cela signifie qu'il est possible de créer des récepteurs nucléaires chimériques, de spécificité différente, en utilisant des domaines fonctionnels provenant de différents récepteurs. Une représentation schématique des différents domaines fonctionnels des récepteurs nucléaires se trouve à la figure 1.3 [5].

#### A) *Domaine A/B*

Cette région est la plus variable en longueur et en séquence parmi les membres de la famille des récepteurs nucléaires. Dans plusieurs cas, cette région contient une fonction d'activation de la transcription autonome, aussi appelée AF-1. Cette fonction est utilisée pour l'activation constitutive et indépendante du ligand du récepteur. La région A/B est aussi une cible pour la phosphorylation provenant de différentes voies de signalisation.

#### B) *Domaine C*

C'est le domaine le plus conservé dans la famille des récepteurs nucléaires. Ce domaine est responsable de la reconnaissance spécifique des HREs et est, en conséquence, aussi appelé domaine de liaison à l'ADN (DBD). Ce domaine sera décrit plus en détails dans la section 1.3.



**Figure 1.3 : Domaines fonctionnels d'un récepteur nucléaire typique.** Chaque domaine est identifié par une lettre, ainsi que les extrémités N et C terminales. La ou les fonction(s) associée(s) à chaque domaine est(sont) inscrite(s) sous le domaine en question.



### *C) Domaine D*

Aussi appelé région charnière, ce domaine n'est pas très conservé parmi les récepteurs nucléaires. Ce domaine sert de charnière entre le DBD et le domaine de liaison au ligand (LBD), permettant ainsi une flexibilité entre ces deux domaines. Dans certain cas, il y a un signal de localisation nucléaire dans la région charnière.

### *D) Domaine E/F*

Ce domaine multifonctionnel est responsable de la liaison du ligand (LBD), de la dimérisation du récepteur, de l'interaction avec les protéines de choc thermique, de l'activation de la transcription dépendante du ligand (fonction AF-2) et, dans certains cas, de la répression de la transcription, cette dernière étant réversible lors de la liaison du ligand.

## **1.2.2 Les éléments de réponses aux hormones**

Les récepteurs nucléaires, étant des facteurs de transcription, doivent lier l'ADN pour réguler la transcription de gènes cibles. Ils reconnaissent des séquences d'ADN habituellement localisées dans la région du promoteur du gène cible. Ces séquences d'ADN sont appelées éléments de réponse aux hormones ou HREs. Deux motifs consensus d'une longueur de six paires de bases ont été identifiés par des études comparatives d'HREs naturels et synthétiques. Les récepteurs nucléaires de classe III (voir le tableau I) reconnaissent préférentiellement la séquence PuGAACA, exception faite du récepteur des oestrogènes. Celui-ci reconnaît le deuxième motif consensus, PuGGTCA, comme le reste de la famille des récepteurs nucléaires. Certains récepteurs nucléaires peuvent reconnaître un motif seul mais dans la plupart des cas, ils reconnaissent deux motifs hexamériques sous forme de dimères. Dans ce dernier cas, les demi-sites peuvent être orientés de différentes façons l'un par rapport à l'autre : ils peuvent former un palindrome, un palindrome inversé (IP) ou une répétition directe (DR). La figure 1.4 montre les différentes orientations possibles des demi-sites ainsi que quelques exemples de récepteurs se liant à chacune des configurations de demi-sites. De plus, les demi-sites sont séparés par une région de longueur variable. Les récepteurs nucléaires de classe III reconnaissent habituellement un palindrome avec une séparation de trois paires de bases. Pour les autres récepteurs



**Figure 1.4 : Configurations possibles des demi-sites liés par les récepteurs nucléaires.** Un demi-site est représenté par une flèche. Le nom de chaque configuration est indiqué sous celle-ci. Quelques exemples de récepteurs nucléaires reconnaissant chaque configuration sont donnés entre parenthèses. N, n'importe quel nucléotide; x, nombre de paires de bases (variable).

nucléaires, le nombre de paires de bases séparant les demi-sites en répétition directe est un facteur pouvant déterminer la sélectivité de la liaison à l'ADN. Par exemple, les HREs de type DR-3, DR-4 et DR-5 sont reconnus respectivement par le VDR, le TR et le RAR (voir le tableau I) [5, 9].

En réalité, les séquences naturelles de HREs démontrent une déviation des motifs consensus. Le tableau II montre quelques exemples de séquences naturelles reconnues par le ER [4]. De plus, il a été démontré que les séquences flanquant les HREs peuvent jouer un rôle dans l'efficacité de la liaison à l'ADN par les récepteurs nucléaires [5]. En somme, les différentes orientations possibles des demi-sites, la variabilité de la longueur de la région séparant les demi-sites et les déviations dans la séquence des motifs consensus donnent lieu à la création d'un vaste répertoire d'HREs. En conséquence, un dimère de récepteurs nucléaires donné peut reconnaître plusieurs types de motifs et s'y lier avec une affinité variable, amenant ainsi l'établissement de réseaux de régulation complexes entre différentes voies de signalisation [1, 5].

### 1.2.3 Modes de dimérisation

La façon dont les récepteurs nucléaires forment des dimères est intimement liée à la configuration des HREs qu'ils lient. Certains récepteurs orphelins, tels que le SF-1, n'ont pas besoin de partenaire pour lier l'ADN. Ils sont capables de lier leur demi-site avec une grande affinité, contrairement aux autres récepteurs [5, 11].

Les récepteurs stéroïdiens (classe III) lient l'ADN presque toujours en formant des homodimères [12]. La liaison au HRE est coopérative; la liaison du premier monomère

**Tableau II : Exemples de séquences naturelles liées par le ER.** L'organisme duquel provient le gène est indiqué entre parenthèses. L'élément de réponse pour le gène de la vitellogénine A2 de *Xenopus* correspond à une séquence consensus. Les demi-sites sont soulignés. Les nucléotides qui diffèrent du consensus sont indiqués à l'aide d'une lettre minuscule. Pour une liste plus complète des éléments de réponses liés par le ER, voir la référence [10].

Gène	Élément de réponse
Vitellogénine A2 ( <i>Xenopus</i> )	5' GTCAGGTCACAGTGACCTGAT 3'
Facteur XII (humain)	5' AGGAGGgCAGCTTGACCaATC 3'
Cathepsine D (humain)	5' GCTGGGcCgGGCTGACCCCGC 3'
c-Fos (humain)	5' GAGccGgCAGCGTGACCCCGG 3'
TGF alpha (humain)	5' GGGGGGTCAGCTgtgCCCCGG 3'

induit la liaison du deuxième monomère et renforce l'interaction du dimère avec l'ADN. Les interfaces de dimérisation se situent dans le DBD (constituée par la boîte D) et le LBD. La configuration du HRE en palindrome force les deux unités de l'homodimère à être orientées de façon symétrique et à avoir un arrangement en « tête-à-tête ». La formation de l'interface de dimérisation, suite à la liaison au HRE, régule la sélection du nombre de paires de bases de la région séparant les deux demi-sites du HRE [5, 11]. En effet, une paire de bases en plus ou en moins résulterait en un mauvais positionnement d'un des DBD face aux paires de bases habituellement contactées par le DBD. Ceci a été démontré avec la structure cristallographique du GRDBD qui lie un élément de réponse ayant quatre paires de bases au lieu de trois entre les deux demi-sites [13].

Certains récepteurs non stéroïdiens peuvent lier l'ADN en tant qu'homodimères mais en général ils forment plutôt des hétérodimères avec le RXR comme partenaire universel. Les interfaces de dimérisation diffèrent selon les récepteurs nucléaires, ce qui permet la liaison à des DRs ayant un nombre de paires de bases variable entre les demi-sites. Les hétérodimères, puisqu'ils lient la plupart du temps des demi-sites en configuration DR, doivent donc avoir un arrangement en « tête-à-queue » [5, 9, 11].

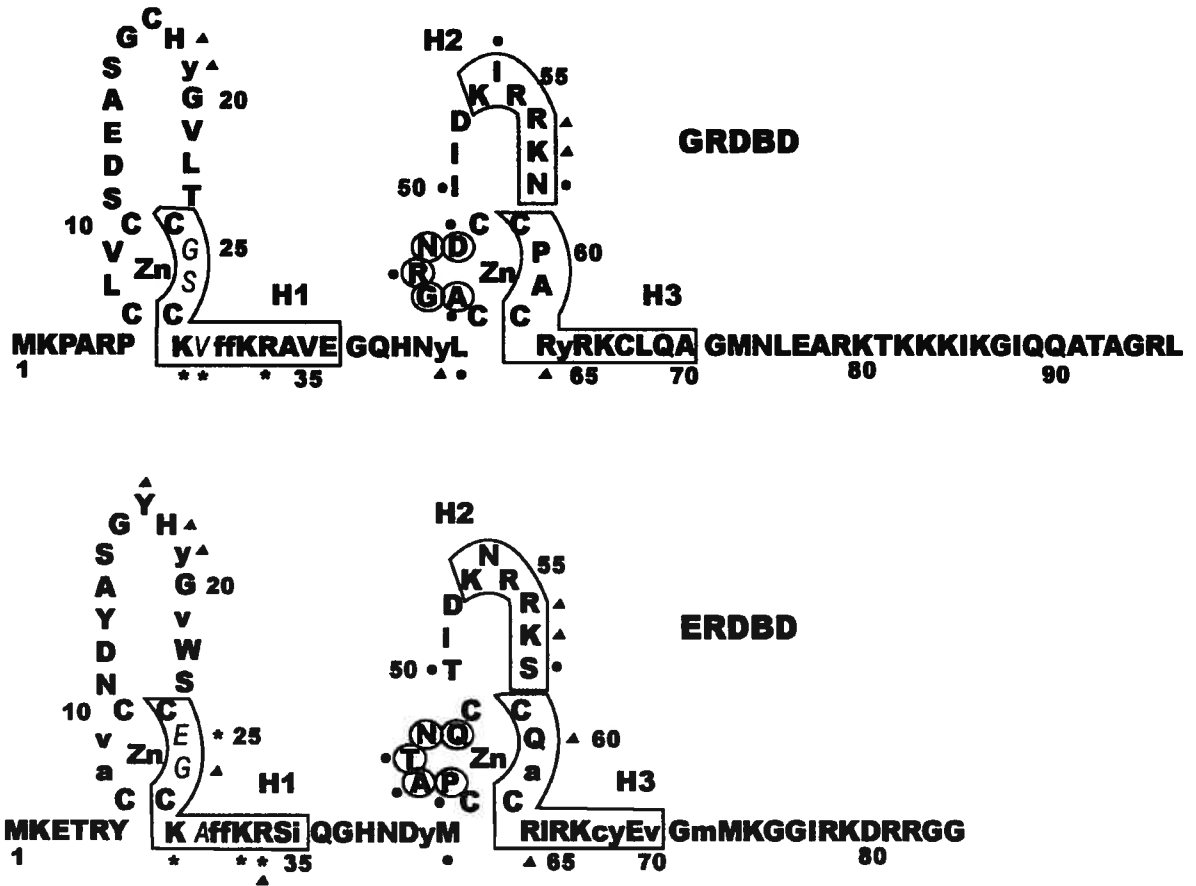
### 1.3 Domaine de liaison à l'ADN

Puisque ce mémoire porte sur l'interaction entre les récepteurs stéroïdiens et l'ADN, la structure et les caractéristiques du domaine de liaison à l'ADN seront décrites en détails dans cette section.

#### 1.3.1 Structures secondaire et tertiaire du DBD

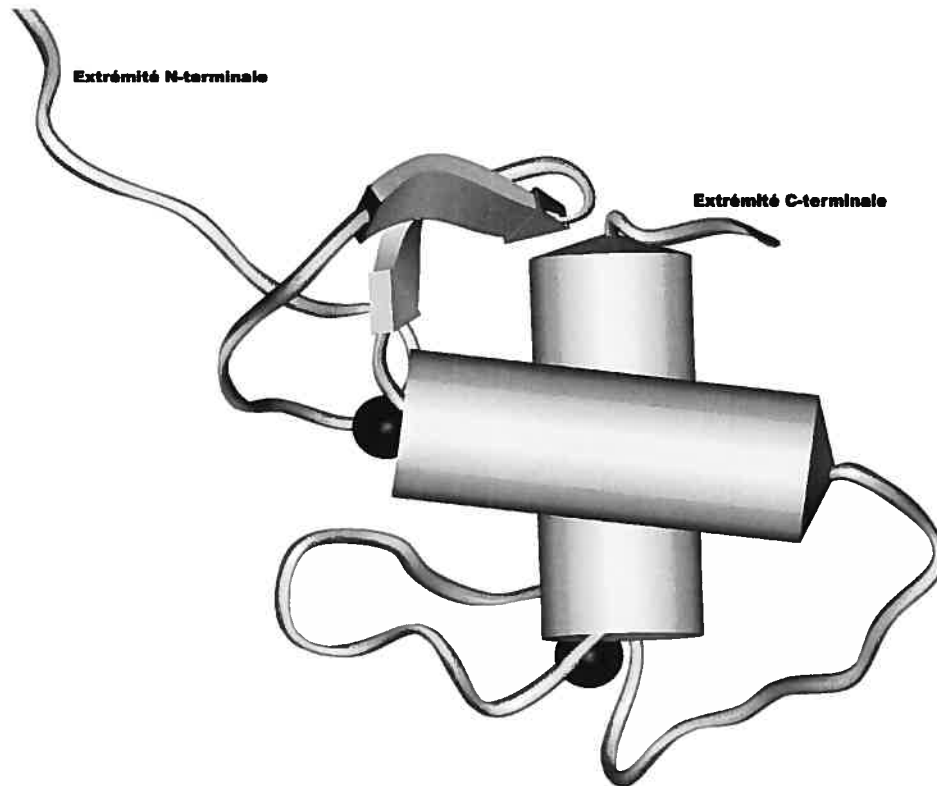
Le DBD correspond à la région C qui est la plus conservée parmi la superfamille des récepteurs nucléaires [14]. Le degré d'homologie varie entre 45% et 91% lorsque les séquences des DBDs des récepteurs stéroïdiens et du récepteur thyroïdien sont comparés entre eux [15]. Les DBDs des récepteurs nucléaires ont donc probablement tous des structures tridimensionnelles similaires sous forme monomérique. Le DBD contient environ 80 acides aminés. La figure 1.5 montre la séquence du GRDBD et du ERDBD, ainsi que leur structure secondaire. Le DBD est composé de deux motifs en doigt de zinc. Les atomes de zinc sont coordonnés par quatre cystéines; on dit donc que ces doigts de zinc sont de type C4. Les deux motifs en doigt de zinc sont séparés par 15 acides aminés. Entre les deux paires de cystéines coordonnant les atomes de zinc, il y a une boucle d'une longueur de 13 acides aminés pour le premier motif et de neuf acides aminés pour le deuxième motif [16]. Le premier motif en doigt de zinc est responsable de la reconnaissance de l'ADN et contient trois résidus responsables de la sélectivité de liaison à l'ADN [17 – 19]. Ces trois résidus forment ce qu'on appelle la boîte P. Leur rôle spécifique sera discuté plus en détails à la section 1.4.1. Le second motif en doigt de zinc est responsable de la dimérisation des DBDs et participe donc à la reconnaissance de l'orientation des demi-sites et de l'espace séparant les demi-sites du HRE [20].

La structure tertiaire du DBD est montrée à la figure 1.6. Les deux motifs en doigt de zinc se replient sur eux-mêmes pour former une seule unité structurale. Dans les deux motifs, une hélice commence après la troisième cystéine coordonnant l'atome de zinc. Le site de liaison de l'atome de zinc se situe donc en N-terminal des hélices  $\alpha$ . Les hélices  $\alpha$  sont d'une longueur variant de 11 à 13 acides aminés et sont fortement amphipathiques. Les deux hélices  $\alpha$  se replient sur elles-mêmes et se croisent à angle droit environ en leur centre.



**Figure 1.5 :** Séquence et structure secondaire du GRDBD et du ERDBD. Les régions hélicales sont encadrées et numérotées de H1 à H3; les acides aminés de la boîte D sont encerclés; les acides aminés de la boîte P sont en caractères italiques non gras; les acides aminés impliqués dans la formation du cœur hydrophobe sont en lettres minuscules. Un astérisque indique un acide aminé faisant un contact direct avec l'ADN; un triangle noir indique un acide aminé faisant un contact avec les groupes phosphates de l'ADN; un cercle noir indique un acide aminé impliqué dans la dimérisation du DBD.

Les acides aminés sur la face hydrophobe des hélices  $\alpha$  forment le cœur hydrophobe de cette unité structurale. On compte de cinq à 15 acides aminés hydrophobes qui sont impliqués dans la formation du cœur hydrophobe et la plupart de ces acides aminés sont conservés dans la famille des récepteurs nucléaires [16].



**Figure 1.6 : Structure tertiaire du DBD des récepteurs stéroïdiens.** Les hélices  $\alpha$  sont représentées par un cylindre; les feuillets  $\beta$  sont représentés par une flèche; les atomes de zinc sont représentés par une sphère noire. Cette figure a été générée à l'aide du logiciel InsightII à partir de la structure du ERDBD [21].

### 1.3.2 Interaction avec l'ADN

Le sujet de ce mémoire porte précisément sur l'interaction entre les récepteurs stéroïdiens et l'ADN. Avant de décrire de façon détaillée l'interface d'interaction entre les DBDs des récepteurs stéroïdiens et l'ADN, quelques éléments généraux concernant les interactions protéines-ADN seront introduits.

#### A) Généralités sur les interactions protéine-ADN

Les protéines reconnaissent une surface avec des caractéristiques chimiques variables en fonction de la séquence d'ADN et des déformations locales dans la structure de l'ADN. Les protéines doivent donc présenter une surface chimique compatible avec la surface de l'ADN pour optimiser leurs interactions avec les paires de bases. De plus, les

protéines liant l'ADN possèdent de nombreux acides aminés chargés positivement qui forment des interactions avec les groupements phosphates de la molécule d'ADN. La portion d'ADN habituellement reconnue par les protéines est composée de quatre à dix paires de bases. L'hélice  $\alpha$  est un élément de structure secondaire communément utilisé pour la liaison à l'ADN. La majorité des contacts protéines-ADN se forment dans le sillon majeur de l'ADN puisque celui-ci est plus exposé que le sillon mineur et parce que chacune des paires de bases peut y être distinguée. Il y a tellement de façons pour une protéine de former une surface chimique complémentaire à une séquence donnée d'ADN qu'il a été jusqu'ici impossible d'établir un code de reconnaissance de l'ADN [22].

Il y a différents types d'interaction possibles entre une protéine et l'ADN. Tout d'abord, mentionnons les ponts hydrogènes (ou ponts H) qui se forment entre un groupe donneur de l'atome d'hydrogène faiblement acide et un atome accepteur porteur d'une paire d'électrons libres. Ce sont des interactions de force moyenne, avec des énergies d'association entre  $-12$  et  $-30 \text{ kJ}\cdot\text{mol}^{-1}$ . Ensuite, il y a les interactions électrostatiques qui se forment entre des groupements chimiques de charges opposées aussi appelées ponts salins. Ce sont des interactions plutôt faibles. Il y a aussi les interactions entre des groupements chimiques électriquement neutres, appelées forces de van der Waals. Ces interactions sont encore plus faibles que les ponts salins [23]. Ces interactions jouent un rôle particulier dans l'interface protéine-ADN puisqu'elles imposent des contraintes stériques sur le type de chaînes latérales et de bases qui peuvent être accommodées dans des positions données. Les interactions de van der Waals, particulièrement entre une thymine et une chaîne latérale hydrophobe d'une protéine, peuvent donc jouer un rôle important dans la sélectivité de la liaison protéine-ADN. Un dernier type de contact observé dans les structures de complexes protéines-ADN est un contact utilisant une molécule d'eau comme trait d'union. Dans ce cas-ci, des ponts H sont formés entre un acide aminé hydrophile de la protéine et une molécule d'eau et celle-ci à son tour peut former des ponts H avec l'ADN. Les molécules d'eau amènent les interactions protéines-ADN à un niveau de complexité encore plus grand. Ce n'est pas encore clair si les molécules d'eau jouent vraiment un rôle dans la spécificité de reconnaissance de l'ADN par les protéines mais il a été proposé que les molécules d'eau serviraient à remplir les espaces laissés vides dans l'interface d'interaction protéine-ADN [22].

### *B) Interaction entre les DBDs et l'ADN*

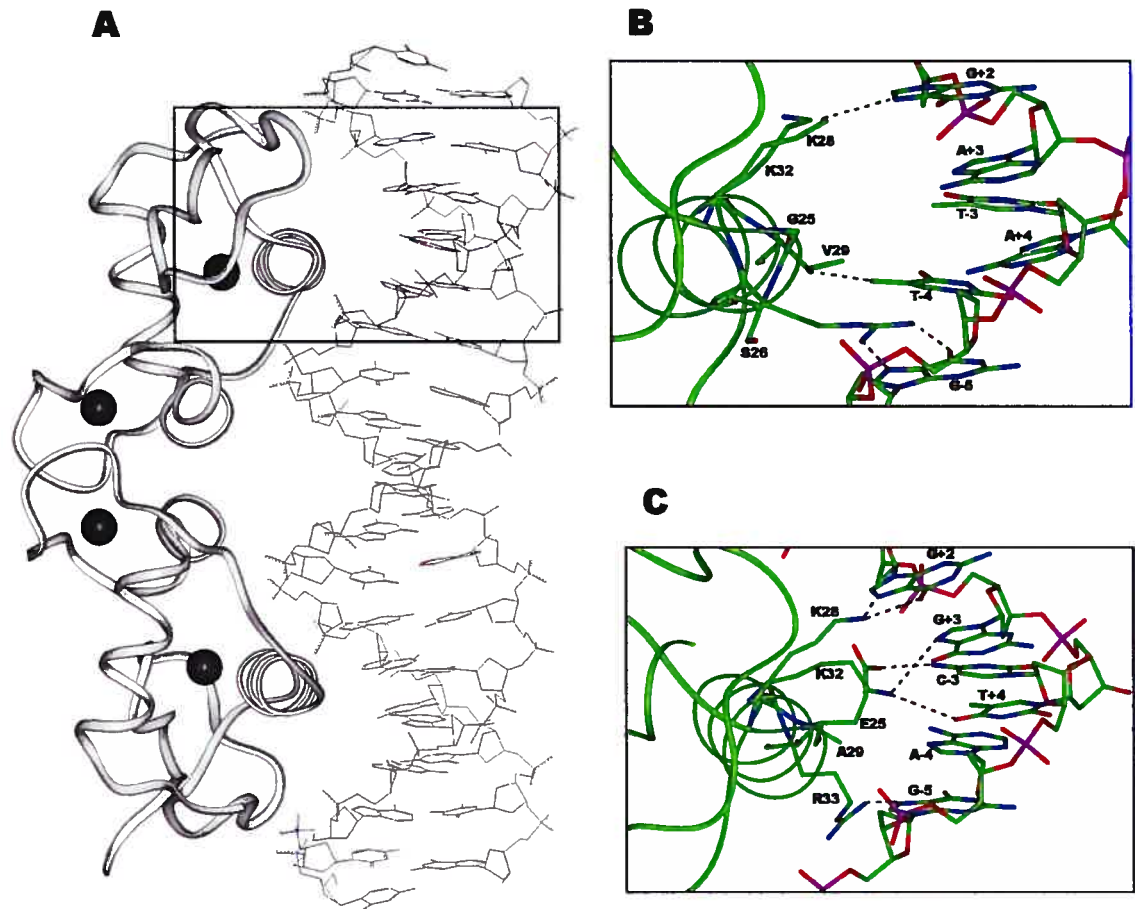
La résolution des structures des récepteurs nucléaires par cristallographie aux rayons X a permis de faire un grand pas dans la compréhension des interactions entre les DBDs de ces récepteurs et l'ADN. Le nombre de complexes résolus ne cesse de croître. Nous ne nous attarderons que sur les deux complexes (GRDBD et ERDBD) qui font l'objet de ce mémoire pour des raisons qui seront expliquées à la section 1.5. La figure 1.7 montre comment les DBDs lient l'ADN et se dimérisent, en plus de montrer le détail de l'interface protéine-ADN. Les deux monomères lient la même face de la molécule d'ADN et les hélices de reconnaissance de l'ADN de chaque DBD se retrouvent dans deux sillons majeurs successifs. Les hélices de reconnaissance de l'ADN (H1) sont positionnées perpendiculairement à l'axe de l'hélice d'ADN. Il n'y a aucun contact dans le sillon mineur. Les deux monomères contactent chacun quatre paires de bases du HRE. De plus, le complexe est symétrique puisque ce sont deux monomères identiques qui lient un palindrome. Pour la suite de l'analyse, un seul monomère sera considéré [13, 21].

Malgré la grande ressemblance des structures des complexes du ERDBD et du GRDBD en interaction avec leur HRE respectif, il y a des différences notables au niveau des contacts avec l'ADN. La figure 1.8 montre de façon schématique les contacts directs entre les chaînes latérales d'acides aminés et les bases de l'ADN. Les contacts avec les groupes phosphates de l'ADN et utilisant les molécules d'eau comme trait d'union ne seront pas considérés dans cette étude pour simplifier le problème. Les différences de contact entre les deux complexes proviennent du fait que la séquence d'ADN reconnue et la séquence de l'hélice de reconnaissance de l'ADN (H1) sont différentes entre les deux récepteurs [13, 21].

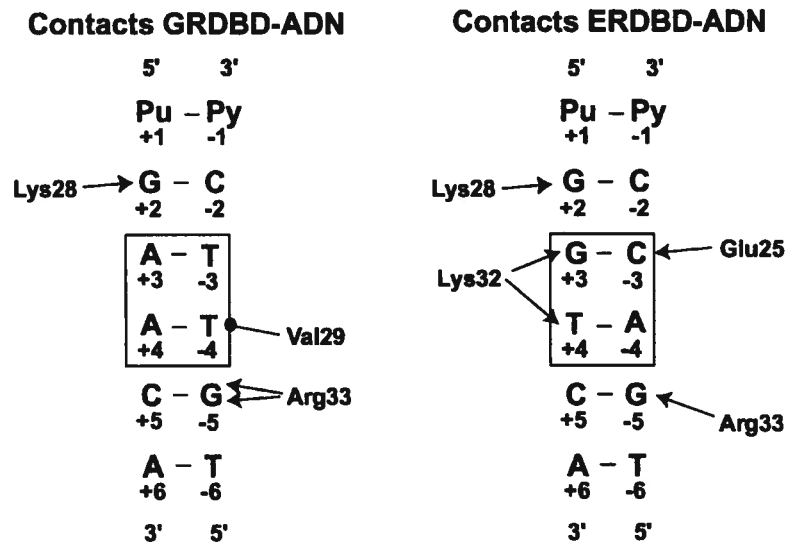
## **1.4 Sélectivité de la liaison à l'ADN**

Une étape importante pour le fonctionnement des récepteurs nucléaires est la liaison à l'ADN, puisque les récepteurs nucléaires doivent reconnaître le bon HRE. Les mécanismes de sélectivité de liaison à l'ADN et les facteurs influençant la liaison à l'ADN seront donc détaillés dans cette section.





**Figure 1.7 : Structure du complexe DBD-HRE et détails de l'interaction entre les DBDs du ER et du GR avec l'ADN.** A : Structure du dimère de DBDs des récepteurs stéroïdiens en interaction avec l'ADN [21]. La chaîne polypeptidique est représentée en ruban; les atomes de zinc sont représentés comme une sphère noire. La région encadrée délimite l'interface d'interaction entre le DBD et l'ADN, détaillée en B et C. B, C : Détails de l'interface d'interaction entre le GRDBD et le GRE (B) et entre le ERDBD et le ERE (C). La chaîne polypeptidique est représentée comme un ruban vert; les atomes de carbone sont en vert, les atomes d'oxygène sont en rouge, les atomes d'azote sont en bleu et les atomes de phosphore sont en magenta; les chaînes latérales des acides aminés d'intérêt sont montrées sous forme de bâtonnets; les atomes de zinc et d'hydrogène ne sont pas représentés pour aérer la figure; les lignes pointillées représentent les ponts H. Pour la numérotation des acides aminés et des nucléotides, se référer aux figures 1.5 et 1.8, respectivement. Cette figure a été générée à l'aide du logiciel InsightII à partir de la structure du ERDBD (A, C) [21] et à partir de la structure du GRDBD (B) [13].



**Figure 1.8 :** Schéma des contacts directs entre les acides aminés du GRDBD et du ERDBD et leur HRE respectif. Les flèches représentent les ponts H et le cercle plein représente une interaction de van der Waals.

#### 1.4.1 La boîte P vs les paires de bases centrales des HREs

Les récepteurs nucléaires peuvent être séparés grossièrement en deux groupes en se basant sur la séquence du demi-site qu'ils reconnaissent. Le premier groupe, représenté par le GR, reconnaît la séquence PuGAACA tandis que le second groupe, représenté par le ER, reconnaît la séquence PuGGTCA [5]. Il y a seulement deux paires de bases qui sont différentes entre les deux demi-sites (numérotées 3 et 4 et soulignées dans les séquences ci-dessus). Il a été démontré que ces paires de bases centrales sont importantes pour la sélectivité de la liaison à l'ADN. Aussi, la plupart des contacts protéine-ADN se forment au niveau de ces paires de bases (figure 1.8) [4].

Malgré la grande conservation dans la séquence du DBD des récepteurs nucléaires, il y a quelques différences dans la séquence en acides aminés entre les récepteurs reconnaissant ces deux types d'éléments. Certains de ces changements se situent au niveau de l'hélice de reconnaissance de l'ADN (H1). Des échanges systématiques d'acides aminés entre le ER et le GR ont démontré que seuls trois acides aminés non conservés faisant partie de cette hélice de reconnaissance de l'ADN (H1) sont importants pour la sélectivité de liaison à l'ADN : il s'agit de Gly25, Ser26 et Val29 pour le GR et Glu25, Gly26 et Ala29

pour le ER. Ces acides aminés forment ce qu'on appelle la boîte P. Lorsque l'identité de ces trois acides aminés dans le ER est changée pour celle des acides aminés dans le GR, le récepteur résultant est capable de lier le HRE spécifique au GR. L'expérience a aussi été réalisée avec un GR contenant les trois acides aminés de la boîte P du ER et la sélectivité de la liaison à l'ADN a aussi été changée [17 – 19].

#### **1.4.2 Autres facteurs jouant un rôle dans la sélectivité de la liaison à l'ADN**

La liaison d'un récepteur nucléaire à un HRE spécifique est un phénomène complexe qui dépend de facteurs autres que la simple interaction entre trois acides aminés et deux paires de bases. De plus, comme discuté à la section 1.2.2, les HREs naturels dans les promoteurs de gènes cibles peuvent différer substantiellement des séquences consensus. La sélectivité de la liaison à un HRE spécifique dépendra donc de l'hydratation du complexe protéine-ADN, de la complémentarité entre les deux surfaces, de la conformation de l'ADN et de réajustements mineurs dans la position des chaînes latérales de certains acides aminés [20]. En ce qui concerne la complémentarité entre les surfaces d'interaction du récepteur nucléaire et de son HRE, il est connu que l'orientation et l'espacement entre les demi-sites déterminent de quelle manière un récepteur nucléaire va se dimériser et par conséquent, quel genre de complexe se liera à l'ADN [24]. Par ailleurs, il est de plus en plus évident que les séquences en nucléotides flanquant les HREs sont aussi des déterminants de la sélectivité de la liaison à l'ADN par les récepteurs nucléaires [25]. Finalement, les molécules d'eau, en créant un réseau complexe de ponts H à l'interface du complexe protéine-ADN, jouent un rôle de plus en plus reconnu dans la sélectivité de la liaison à l'ADN [26].

### **1.5 Code de reconnaissance de l'ADN**

De plus en plus de complexes protéines-ADN sont résolus par cristallographie aux rayons X et il y a une forte accumulation de données expérimentales concernant la sélectivité de liaison à l'ADN par divers groupes de protéines. Ces résultats ont poussé les chercheurs à tenter de prédire la séquence en ADN qui pourrait être reconnue par une protéine ayant une séquence en acides aminés particulière dans son hélice de

reconnaissance à l'ADN. Un tel code de reconnaissance de l'ADN a été publié en 1994 par Suzuki et Yagi pour les protéines à doigts de zinc de type C4, ce qui inclut les récepteurs nucléaires (voir la figure 2.2) [27]. Le code de reconnaissance de l'ADN se base sur deux types de règles : les règles chimiques et les règles stéréochimiques. Les règles chimiques décrivent les appariements possibles entre les acides aminés et les bases par liaisons hydrogènes ou interactions hydrophobes. De plus, la taille des acides aminés est aussi prise en compte dans ces règles. Les acides aminés sont classés en fonction de leur taille dans un de ces quatre groupes généraux : petit, moyen, large ou aromatique. En effet, un acide aminé situé à une position fixe sur une hélice de reconnaissance de l'ADN peut rejoindre une base dans le sillon majeur de l'ADN située plus ou moins loin en fonction de la taille de sa taille. Les règles stéréochimiques sont spécifiques à chaque groupe de protéines car celles-ci dépendent de la position de l'hélice de reconnaissance de l'ADN dans le sillon majeur de l'ADN. Les règles stéréochimiques sont présentées en partie pour les récepteurs nucléaires à la figure 2.2 B et C. Un tel code pourrait être utilisé pour la création d'un récepteur nucléaire mutant capable de reconnaître une séquence d'ADN prédéterminée.

## 1.6 Questions abordées dans ce mémoire

Malgré l'abondance de données expérimentales et structurales, le mécanisme de sélectivité de liaison à l'ADN utilisé par les récepteurs nucléaires n'est pas encore tout à fait compris. Plusieurs paramètres décrits dans les sections précédentes commencent à peine à être intégrés dans les études actuelles. Ce projet de recherche a pour but de découvrir ces paramètres structuraux encore non couverts par la littérature et d'explorer comment ils influencent la sélectivité de la liaison à l'ADN des récepteurs nucléaires.

Un aspect intéressant à noter à propos des récepteurs nucléaires est la composition de leur boîte P respective et de la séquence du demi-site qu'ils reconnaissent. À la section 1.4.1, les récepteurs nucléaires ont pu être séparés en deux groupes selon la séquence du demi-site qu'ils lient. Le premier groupe de récepteurs nucléaires, reconnaissant la séquence PuGAACA, ont tous une boîte P dont la composition est  $G_{25}S_{26}V_{29}$  comme pour le GR. Les récepteurs nucléaires faisant partie de ce groupe sont : AR, MR et PR. Le reste de la famille des récepteurs nucléaires, incluant le ER, font partie du second groupe reconnaissant la séquence PuGGTCA. Ils ont tous une boîte P dont la composition est

exactement ou similaire à E<sub>25</sub>G<sub>26</sub>A<sub>29</sub>. C'est pour cette raison que seules les structures du GRDBD et du ERDBD ont été utilisées lors de cette étude puisque ces récepteurs sont des prototypes de chacun des deux sous-groupes de récepteurs nucléaires. Il y a donc une forte corrélation entre la composition de la boîte P des récepteurs nucléaires et la séquence du demi-site qu'ils reconnaissent. Il est donc possible de suggérer, en accord avec les études publiées à ce sujet [17 – 19], que la sélectivité de la liaison à l'ADN dépend de la composition de la boîte P. Pourtant, les structures tridimensionnelles du GRDBD [13] et du ERDBD [21] en interaction avec leur HRE respectif démontrent qu'un seul acide aminé de la boîte P fait un contact direct avec l'ADN dans chacun de ces complexes (figure 1.8). De plus, les acides aminés de la boîte P contactant directement l'ADN ne sont pas équivalents dans les deux récepteurs : dans le cas du GRDBD, il s'agit de Val29 (deuxième acide aminé de la boîte P) qui contacte T-4 alors que dans le cas du ERDBD, il s'agit de Glu25 (premier acide aminé de la boîte P) qui contacte C-3.

Un aspect intéressant à propos de la composition de la boîte P des récepteurs nucléaires est la présence d'une glycine soit à la première ou à la deuxième position de la boîte P. La caractéristique principale d'une glycine est l'absence de chaîne latérale donc il est difficile de concevoir comment cet acide aminé peut être un important déterminant de la sélectivité de la liaison à l'ADN. Il a été suggéré que cette glycine pourrait être impliquée soit dans le positionnement correct de H1 dans le sillon majeur de l'ADN ou soit comme un initiateur de la formation de l'hélice  $\alpha$  [28]. Avec la résolution des structures tridimensionnelles du GRDBD [13] et du ERDBD [21], il a été clarifié que le rôle de la glycine n'était pas celui d'un initiateur d'une hélice  $\alpha$  puisque H1 débute à l'acide aminé précédent la boîte P. De plus, ce n'était pas clair comment une glycine pourrait initier la formation d'une hélice  $\alpha$ . Par ailleurs, il est difficile d'établir en quoi la formation d'une hélice  $\alpha$  soit en première ou en deuxième position de la boîte P pourrait être un déterminant de la sélectivité de liaison à l'ADN. Par contre, il est raisonnable de suggérer que le bon positionnement de H1 dans le sillon majeur de l'ADN pourrait être un facteur déterminant de la sélectivité de liaison à l'ADN et que la boîte P, incluant la glycine en première ou deuxième position, serait responsable de ce mécanisme de sélection. Pourtant, cet aspect a été mentionné dans certaines études [16, 28] sans pour autant être exploré en détail. De plus, il a été récemment démontré que le contexte structural d'une protéine en interaction avec l'ADN était crucial pour la prédiction des interactions protéine-ADN [29].

Dans ce mémoire, nous nous efforcerons d'élucider le mécanisme moléculaire de la sélectivité de la liaison à l'ADN par les récepteurs stéroïdiens. Pour ce faire, nous testerons dans un premier temps les prédictions du code de reconnaissance de l'ADN de Suzuki en faisant des remplacements systématiques dans les paires de bases du HRE. Ensuite, nous procéderons en analysant les structures tridimensionnelles du GRDBD et du ERDBD en interaction avec leur HRE respectif dans le but d'éclaircir le rôle du positionnement de H1 dans le sillon majeur. Un modèle sera proposé en dernier lieu qui intégrera ce nouveau facteur déterminant de la sélectivité de la liaison à l'ADN et qui redéfinira les rôles de chacun des acides aminés interagissant avec l'ADN.

**CHAPITRE 2 : RÉSULTATS, 1<sup>ÈRE</sup> PARTIE (ARTICLE)**

**RATIONAL DESIGN OF ESTROGEN RECEPTOR MUTANTS WITH ALTERED  
DNA BINDING SPECIFICITY**

Denis Nguyen, Genevieve Pesant, Étienne Rouault, Julie Deschênes, Walter Rocha, Samuel Chagnon , Sergey V. Steinberg and Sylvie Mader<sup>#</sup>

Département de Biochimie, Université de Montréal, C.P. 6128 Succursale Centre Ville,  
Montréal Qc H3C 3J7 Canada.

<sup>#</sup> To whom correspondance should be addressed at : Département de Biochimie, Faculté de Médecine, Université de Montréal, C.P.6128 Succursale Centre Ville, Montréal Qc H3C 3J7 Canada.

Tel: 514-3436111#5173, Fax: 514-3432210, [REDACTED]



## **2.1 Contribution des auteurs**

Denis Nguyen : responsable de toutes les figures des manipulations expérimentales.

Geneviève Pesant : analyse des résultats expérimentaux en utilisant la technique de modélisation moléculaire.

Étienne Rouault, Julie Deschênes, Walter Rocha, Samuel Chagnon : élaboration des constructions ayant servies aux manipulations expérimentales.

Sergey V. Steinberg : superviseur de recherche de Geneviève Pesant.

Sylvie Mader : superviseur de recherche des autres auteurs, co-superviseur de recherche de Geneviève Pesant, chercheur principal du projet de recherche.

## 2.2 Summary

Although rational design of sequence-specific DNA binding domains (DBDs) has been achieved successfully in the case of the C2-H2 class of zinc fingers, it remains unclear to which extent the C4 zinc fingers found in nuclear receptors can be tailored to bind novel response elements in a transcriptionally-productive manner. Nuclear receptors bind as dimers with high selectivity to response elements containing variations of two basic 6 base pair motifs, through specific contacts established by a few critical amino acids. The two central bases of these motifs are critical for differential recognition of response elements by different steroid receptors. The crystal structure of the estrogen receptor (ER) in complex with its cognate estrogen response element (ERE) indicates a complex mode of recognition of these two central base pairs in the PuGGTCA motifs. Two charged amino acids, Glu 203 and Lys 210, establish specific contacts each with the two central bases, but on the opposite DNA strands. In this study, mutagenesis analysis indicated that both these amino acids contribute to binding affinity on the consensus ERE. Removal of the electric charge and steric constraints associated with the presence of Lys 210 was found to be required for full loss of parental DNA binding specificity and/or acquired recognition of novel sequences by Glu 203 mutants. Even in the absence of a charged residue at position 210, all changes of specificity predicted from the nature of residue 203 were not obtained. However, the double mutation E203R-K210A generated a mutant ER that bound and transactivated palindromes of PuGCTCA motifs, but not consensus EREs, demonstrating the feasibility of designing C4 zinc finger mutants with novel DNA binding specificity in the two central base pairs of palindromic response elements.

## 2.3 Introduction

Nuclear receptors form a superfamily of ligand-inducible transcription factors that is characterized by two conserved domains, the DNA binding domain (DBD) composed of two C4 type zinc fingers, and the ligand binding domain (LBD), which also contains a dimerization interface (1,2). Nuclear receptors can bind DNA as homo- and/or heterodimers, and recognize response elements arranged as direct repeats, palindromes or inverted palindromes of conserved motifs (3-5). Each motif is bound by the DBD of a

single monomer, the two zinc fingers of the DBD combining into a single structural fold with a DNA recognition helix and variable dimerization interfaces (6,7).

Estrogen receptors interact specifically with palindromes of PuGGTCA motifs with a three base pair spacer (8-10) also called estrogen response elements (EREs). Perfect or imperfect EREs have been found in several estrogen responsive promoters (10-12). In addition, high affinity EREs characterized at up to 10 kb distance from the transcriptional start site of estrogen regulated genes are bound by ER $\alpha$  *in vivo* (13). Other steroid receptors also bind palindromes with a three base pairs spacer, but the repeated motifs are PuGNACA sequences (GREs, (14,15)). Non-steroid receptors also recognize PuGGTCA or related motifs, but these motifs are arranged as direct repeats or everted repeats with variable spacing. The affinity and selectivity of nuclear receptors for single PuGGTCA motifs is generally low, but can be increased due to receptor-specific recognition of additional 5' flanking residues (5,12). Thus nuclear receptors have achieved selectivity in DNA recognition while interacting with only two different types of basic motifs. The fact that few variations have been observed in the base-contacting amino acids of all the mammalian nuclear receptors suggests that this type of protein-DNA recognition has been conserved throughout evolution, possibly because it affords the most favourable combination of affinity and selectivity. Interestingly however, nuclear receptor homologs identified in *C. elegans* offer considerably higher variety in the amino acid composition in the DNA binding helix, although it is not clear at this point whether all these proteins actually bind DNA. Numerous mutations have also been described in the DNA binding domain of some nuclear receptors such as VDR and AR, but few changes in DNA binding patterns have been reported for these mutant receptors (16).

The first clues to the molecular basis of specific DNA recognition by steroid receptors were provided by mutagenesis experiments of the estrogen and glucocorticoid receptors (GRs), which led to identification of the amino acids that are responsible for discrimination between the two types of recognition motifs in the palindromic motifs of these receptors. Exchanging three amino acids in the ER DBD for the corresponding ones in the GR DBD resulted in a receptor mutant capable of transactivating glucocorticoid target promoters (17). The converse experiment also demonstrated that amino acids at the same three positions (P box) in the GR were crucial for discrimination of glucocorticoid versus estrogen response elements (18,19). In addition, a loop in the second zinc finger is

responsible for specific recognition of two motifs arranged as palindromes with 3 bp spacing (18-20). Crystallographic analyses of complexes between the ER or GR DNA binding domains and their response elements have uncovered the amino acid – base pair interactions responsible for the differential DNA binding selectivity of these two receptors (6,7,21,22). Two residues of the P box, Val at the third position in the GR and Glu at the first position in the ER (Glu 203), contact the central discriminating bases in the ERE and GRE motifs. In addition, an invariant Lys residue located farther downstream in the DNA recognition helix (Lys 210 in ER $\alpha$ ) binds the two discriminating bases on the opposite strand with respect to Glu 203 in the ER $\alpha$ -ERE complex, but does not participate in contacts in the GR-GRE complex. Other bases common to the ERE and GRE are contacted by amino acids conserved in ER and GR (Lys 206 and Arg 211 in ER $\alpha$ ). In addition, these interactions are buttressed by a complex network of amino acid-amino acid interactions and amino acid-phosphate interactions.

The specificity of response motif recognition by steroid receptors is thus determined by a small number of specific interactions established by 4 to 5 amino acids. As a consequence, it may be expected that changing the identity of interacting amino acids in the DNA recognition helix would alter the selectivity of receptor/DNA interaction. A simple model for amino acid-base interactions within the structural framework of the steroid receptor DNA binding domain has been proposed (23). This model relies on chemical rules for possible pairing of amino acid side chains and DNA bases through hydrogen bonding or hydrophobic interactions, and also incorporates stereochemical constraints specific to steroid receptors, based on the position of the DNA recognition helix with respect to the major groove of the DNA. Small, medium or large chains may thus be preferred depending on the position of the interacting amino acids in the DNA recognition helix with respect to the bases. Study of a spontaneous mutation in the first amino acid of the P box of the androgen receptor has revealed changes in DNA binding specificity compatible with the predictions of this model (16). Replacement of Gly at the first P box position in the AR DBD by Arg resulted in a receptor that could only bind a subset of the PuGNACA motifs normally bound by this receptor, and this according to the chemical preferences of Arg residue with respect to the adjacent base at the third position of the motif. However, it remains unclear whether novel types of DNA specificities can be achieved through rational design as described for the C2H2 type of zinc fingers (24-26), especially in the case of the

ERs, which appear to have more complex determinants of motif recognition than other steroid receptors.

Here we have investigated further the molecular basis for the DNA binding specificity of estrogen receptors to sequences derived from consensus EREs by introduction of systematic base replacements at each position of the motif. These studies are of interest not only to better understand how the primary sequence of the ERE affects ER binding, information is necessary for the identification of estrogen response elements in genomic sequences (13), but also to determine whether and how the specificity of homodimeric steroid receptors can be altered to achieve specific recognition of novel palindromic response elements.

## **2.4 Materials and methods**

### **2.4.1 Plasmids**

The bacterial expression vector pET31-HE81 containing the ER $\alpha$  DBD has been described previously (20). The bacterial expression vector pET31-ER $\beta$ -DBD was constructed by PCR amplification of a cDNA fragment corresponding to amino acids 140 to 246 of ER $\beta$  and subcloning between the KpnI and XhoI sites of a pET31, a pET3 vector derivative modified by insertion of a linker containing the KpnI and XhoI sites (20). All ER $\alpha$  DBDs mutants were constructed by insertion between the KpnI and XhoI sites of the pET31 vector of a fragment obtained by site-directed mutagenesis of pET31-HE81 using PCR amplification.

The wild-type ER $\alpha$  and ER $\beta$  expression vectors pSG5-HEG0, pCMV-SPORT-ER $\beta$  have been described previously (13,27). The expression vector pSG1-ER $\alpha$ 203G has also been described previously and corresponds to HE84 (17). Expression vectors for ER $\alpha$ E203A, ER $\alpha$ E203N, ER $\alpha$ E203H, ER $\alpha$ E203R, ER $\alpha$ K210A, ER $\alpha$ E203R-K210A have been constructed by substitution of the fragment between the KpnI and XhoI sites of pSG1-HE80 (20) by a fragment containing the mutation obtained by site-directed mutagenesis using PCR amplification.

The tk-CAT reporter plasmids containing one copy of the consensus ERE or of palindromes containing base replacements (pAT-tkCAT, pCT-tkCAT) were derived from

the pBL-CAT8+ reporter vector (28) by insertion of double stranded oligonucleotides containing the response elements flanked with BamHI-BglII sites at the BamHI site upstream of the thymidine kinase promoter. The TATA-CAT reporter vectors containing two copies of the consensus ERE or of the CT palindrome were prepared by substitution of the three EREs in pERE3-TATA-CAT (29) by double stranded oligonucleotides containing two tandem response elements spaced by 31 bp and flanked by BamHI and BglII sites.

#### **2.4.2 E. coli expression of the ER $\alpha$ and ER $\beta$ DNA binding domain and derivatives thereof**

E. coli BL21 DE3 cells were transformed with pET31 expression vectors containing the cDNAs for the DNA binding domains of ER $\alpha$ , of mutants of ER $\alpha$ , or ER $\beta$  and expression was induced in exponentially growing cultures by IPTG (0.5 mM final) for 2 h. Whole bacterial extracts were prepared by sonication in extraction buffer (Tris HCl pH 7.4, 25 mM; EDTA pH 8.0, 0.1 mM; NaCl 400 mM; glycerol 10%; DTT 1 mM; PMSF 10 mM and protease inhibitors) followed by centrifugation (at 10 000g for 30 min) of bacteria.

To determine the levels of expression of ER DBDs, aliquots (1 ml) were taken from each culture before induction with IPTG. Bacteria were isolated by centrifugation and resuspended in M9 medium containing each amino acid except methionine and cysteine (0.01% weight/volume each). Rifampicine was added (200  $\mu$ g/ml final) to inhibit bacterial RNA polymerase and expression of the T7 polymerase was induced with IPTG (0.5 mM final) for 30 min. [<sup>35</sup>S]-methionine was then added and bacteria were further incubated at 37°C for 5 min, collected by centrifugation, resuspended in Laemmli buffer, and boiled for 5 min. Labelled proteins were analysed by electrophoresis on 12% polyacrylamide-SDS gel and visualized by fluorography.

#### **2.4.3 Immunoblot analyses**

COS-1 cells were transiently transfected with expression vectors for the wild-type ER $\alpha$  or mutants thereof and lysed in extraction buffer (Tris HCl pH 7.6, 20 mM; glycerol 20%; KCl 400 mM; DTT 1 mM; PMSF 10 mM and proteases inhibitors). Protein concentration in whole cell extracts was estimated using a Bradford assay. Protein extracts

(5 µg) in Laemmli buffer were separated by electrophoresis on 8% polyacrylamide-SDS gel, transferred onto a PVDF membrane. Blots were incubated in blocking solution (TBS 1X, Tween 20 0.05%, BSA 3%) for 20 min and probed with anti-ER $\alpha$  mouse monoclonal antibody B10 (obtained from Prof. P. Chambon, dilution 1:5000). Membranes were then washed and incubated with a horseradish peroxidase-coupled secondary antibody and visualized with an ECL detection kit (NEN Life Science Products).

#### 2.4.4 Gel shift assays

Whole cell extracts from COS-1 cells transiently transfected with ER expression vectors (15 µg per 9 cm dish) were diluted to a final KCl concentration of 80 to 125 mM. Samples were preincubated with 2 µg poly(dIdC) for 15 min on ice before addition of [ $^{32}$ P]-labelled double stranded oligonucleotide probes (50 000 cpm per sample). Reactions were incubated at 25°C for 15 min then loading buffer (0.1% bromophenol blue, 60% glycerol) was added (16% V/V). Complexes were separated by electrophoresis on 5% polyacrylamide gels in 0.25 or 0.5 X TBE (45 mM Tris-HCl, 45 mM boric acid and 1 mM EDTA) and visualized by autoradiography.

Gels shift assays with whole bacterial extracts containing the ER $\alpha$  or Er $\beta$  DBD were performed as described above except that extracts were diluted to a final NaCl concentration of 80 mM and 7% polyacrylamide gels in 0.25 X TBE were used to separate the complexes.

For competition assays, 60 µl reactions were preincubated on ice for 15 min before addition of 50 fmol [ $^{32}$ P]-labelled oligonucleotide probes (100 000 cpm per sample). Samples were incubated at 25°C for 15 min and loading buffer (0.1% bromophenol blue, 60% glycerol) was added (16% V/V). 50 pmol of unlabelled oligonucleotide were added to each samples and 5 µl aliquots of each reaction mix were taken immediately or 5, 10, 15, 20, 25, 30 and 35 min later and loaded on continuously running 7% polyacrilamide gels in 0.25 or 0.5 X TBE (45 mM Tris-HCl, 45 mM boric acid and 1 mM EDTA). Controls performed without adding unlabelled oligonucleotide or by adding unlabelled oligonucleotide premixed with the [ $^{32}$ P]-labelled oligonucleotide probes were performed. Complexes were quantified using a Phosphorimaging screen and the Quantity One software from Bio-Rad.

### 2.4.5 CAT assays

HeLa cells were maintained in DMEM supplemented with 5% fetal bovine serum (FBS). Cells were switched 3 days before transient transfection to medium without phenol red containing 5% FBS pretreated with activated charcoal to remove traces of hormones. HeLa cells were electroporated ( $5 \times 10^6$  cells, 0.24 kV, 950  $\mu$ F in a Bio-Rad Gene Pulser II apparatus) with varying amounts of expression vectors for wt ER $\alpha$  or for different ER $\alpha$  mutants, 2  $\mu$ g pCMV- $\beta$ Gal, 2  $\mu$ g of tk-CAT reporter vectors containing single copies of different palindromic response elements. DNA mixes were completed to 40  $\mu$ g with salmon sperm DNA in a final volume of 50  $\mu$ l of NaCl 210 mM. Cells were treated or not with estradiol (25 nM) for 40 h. Cells were then harvested and extracts were prepared by freeze thawing in Tris HCl pH8.0 (0.25 M). CAT assays were performed and standardized over  $\beta$ -galactosidase activity. For transactivation experiments with TATA-CAT reporter vectors containing two copies of the consensus ERE or of the CT element, electroporation was performed essentially as described above except that twice as many cells and double volumes of DNA mixes were used.

### 2.4.6 Modeling

Modeling was performed interactively, using InsightII/Discover package (version 2000, Accelrys Inc., San Diego, CA). The X-ray structure of the ER DBD bound to DNA (ref. Schwabe et al. (1993) Cell) was used as a starting conformation. Each model was submitted to unrestrained energy minimization using the AMBER forcefield (30) until an energy minimum was reached. The presence or absence of particular pair-wise amino acid – base interactions in the final structure was treated as a possibility or impossibility to form a particular interaction in a given structural context. Distance measurements between atoms were done with InsightII tools. Visualizations were done in a Silicon Graphics O2 computer.

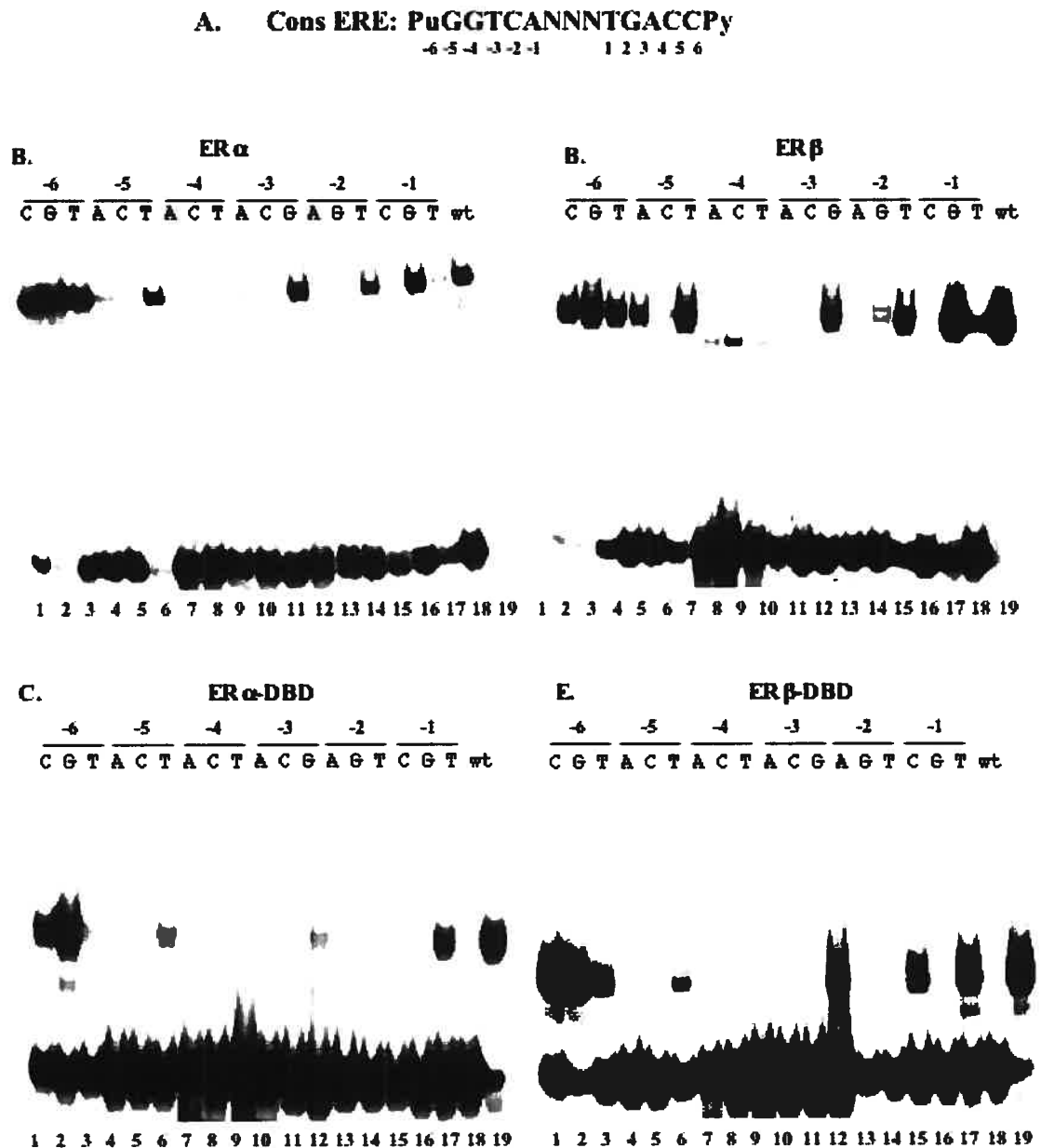


## 2.5 Results

### 2.5.1 The patterns of binding of both ERs to elements containing base replacements in both motifs are similar to each other and to those of the isolated DBDs.

Estrogen receptors bind with high affinity the consensus palindromic EREs consisting of two PuGGTCA motifs separated by a three base pair spacer (figure 2.1 A). Although natural response elements are often imperfect palindromes (10,12), base pair replacements usually result in a loss of affinity (see ref. (22) for a structural study of the impact of a single nucleotide replacement on protein-DNA interface and binding affinity). In order to analyze the molecular basis of the specific recognition of palindromic EREs by estrogen receptors, we initially examined the effect of symmetric substitutions at each position of the consensus ERE on binding by estrogen receptors. A panel of probes derived from the perfect *Xenopus* Vitellogenin A2 ERE were used in gel shift assays with either whole cell extracts of COS-1 cells transiently transfected with expression vectors for wt ER $\alpha$  or ER $\beta$  (figure 2.1 B, C) or bacterial extracts expressing the ER $\alpha$  or ER $\beta$  DNA binding domain (ER $\alpha$  DBD and ER $\beta$  DBD, figure 2.1D, E). The wild type ERE was bound by either the full-length ER $\alpha$  receptor or the corresponding DBD with the highest affinity (note the higher degree of free probe depletion with wt ERE, figure 2.1B-E, lanes 19). Consistent with our previous observations that the DNA binding domains of steroid receptors are sufficient for selective response element recognition (20), ER DBDs selectively bound the array of probes in a pattern similar to that of the full length receptors, although all probes were less efficiently bound with isolated DBDs. This is consistent with a loss of affinity, but not specificity, resulting from the absence of the strong dimerization interface in the ligand binding domain (31).

All replacements in both arms of the ERE reduced binding to the same extent for ER $\alpha$  and ER $\beta$  (compare panels B and D, and C and E from figure 2.1), which share a high degree of conservation in their DNA binding domains. Although the expression levels of the full length ER $\alpha$  and ER $\beta$  transiently transfected in COS cells could not be compared, the DBDs were expressed to similar levels as assessed by [<sup>35</sup>S]Met incorporation (data not shown). Binding levels to our panel of probes appear similar for the two receptor DBDs, an observation supported by their similar off rate curves (data not shown).

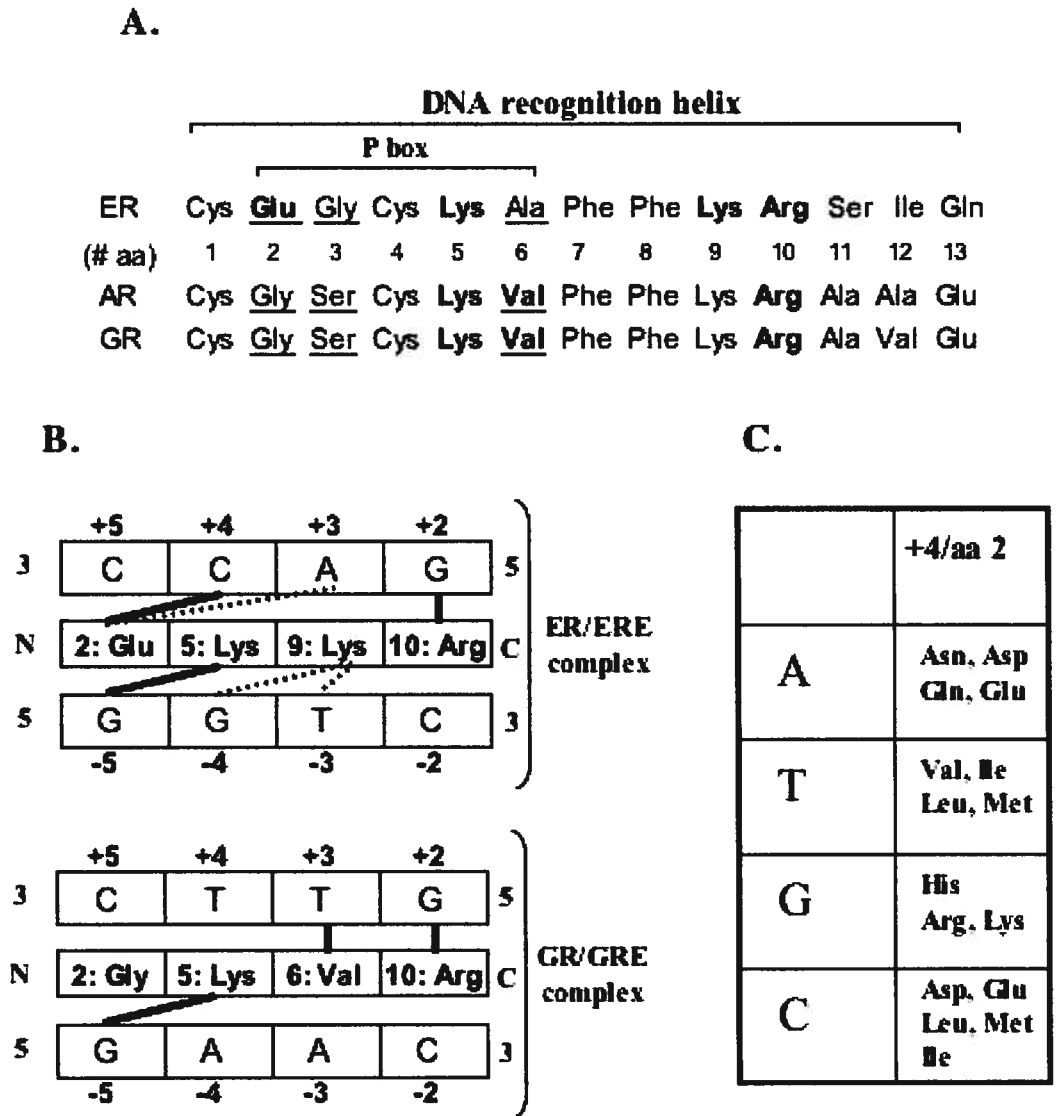


**Figure 2.1 : Effect of replacements in both arms of the ERE palindrome on complex formation with full length ERs and isolated DBDs. A :** Sequence of the consensus ERE and base numbering used in this study. **B-C :** Gel shift assays performed with whole cell extracts of HeLa cells transiently transfected with expression vectors for ER $\alpha$  (**B**) or ER $\beta$  (**C**), and a panel of probes corresponding to all base possibilities introduced at each position of the response elements in both arms of the palindrome. **D-E :** Gel shift assays performed with bacterial extracts containing the DNA binding domains of ER $\alpha$  (**B**) or ER $\beta$  (**C**), using the same panel of probes as in **B-C**.

### 2.5.2 Point mutations in the consensus ERE destabilize interaction with ERs according to chemical compatibility and steric constraints with interacting amino acids

The DNA recognition helix formed by the C-terminal part of the first zinc finger contains several basic amino acids involved in contacts with bases in the target motifs (figure 2.2A, residues in bold). Nucleotides G (-5) and G (+2) interact with residues at position 5 (Lys 206) and 10 (Arg 211) of the DNA recognition helix, respectively (figure 2.2B, bold lines). In addition, Lys 210 (position 9 in the DNA recognition helix) interacts with both G (-4) and T (-3) through direct and water-mediated contacts (figure 2.2B, dotted lines). Thus, basic residues are involved in recognition of all positions in the ERE except -/+6 and -/+1, which are not directly contacted in the crystal structure although both display a preference for purines (figure 2.1B-E, lanes 1-3 and 16-18). These direct interactions involving basic amino acids conform to general chemical rules, with G and T, which present only negatively charged groups in the major groove, being preferred over A, which present both a positively and a negatively charged group, and C residues, which contain only a positively charged group leading to unfavourable electrostatic interactions. Accordingly, replacement of G in a single motif at position -4 or -5 was sufficient to abolish binding. Of note however, the order of the preferred bases is not identical for each contacting basic amino acid. Most noticeably, replacement in a single motif of G at position +2 by T was sufficient to abolish binding, whereas a faint complex was detected with C (figure 2.3A). Molecular modeling suggests that lack of binding to T at position +2 results from steric hindrance due to the methyl group of T+2, which prevents productive interaction between the amines of Arg 211 and O4 of T+2 (figure 2.3B).

The pattern of recognition of bound response elements carrying replacements at positions -4/-3 also slightly differed from that predicted from charge preference due to interaction with basic amino acid Lys 210. We have examined the potential role of another amino acid, Glu 203, which was shown in the crystal structure to interact also with bases at the same positions in the ERE, but on the other strand (7), in this pattern of probe recognition. Glu 203 has an opposite and complementary type of chemical selectivity for bases compared to Lys 210, but the complex mode of recognition of the two central base pairs by two amino acids may be expected to result in increased specificity due to steric



**Figure 2.2 : Model for the selectivity of steroid receptor/DNA interactions.** A. DNA recognition helix of the estrogen (ER), glucocorticoid (GR) and androgen (AR) receptors. B. Models of the amino acid - base interactions underlying specific recognition of the estrogen (top) or glucocorticoid (bottom) response element by their cognate steroid receptors. Interactions considered in the model proposed by Suzuki and Yagi (23) are in bold, while interactions described in the ER-ERE crystal structure (7) but not in the model are in dashed lines. C. Predicted interactions between amino acid replacement at position 2 of the DNA recognition helix and bases at position +4 (from (23)).

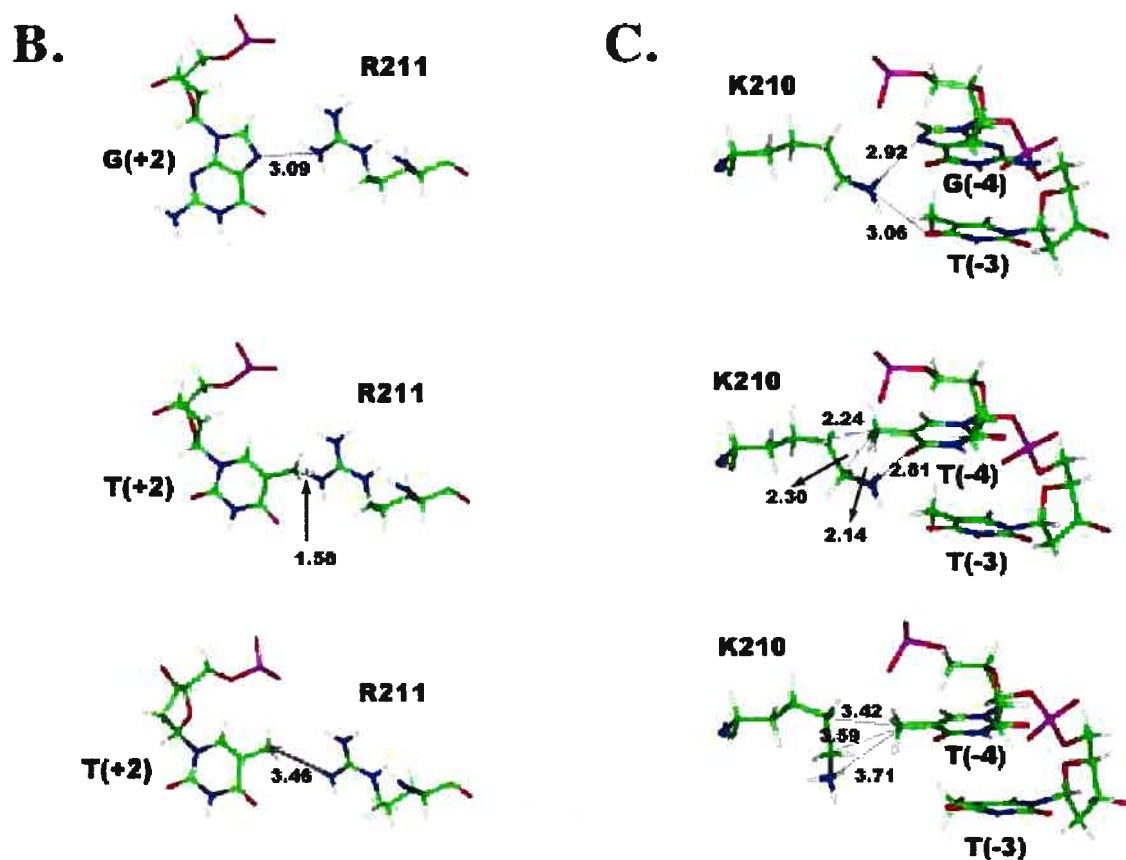
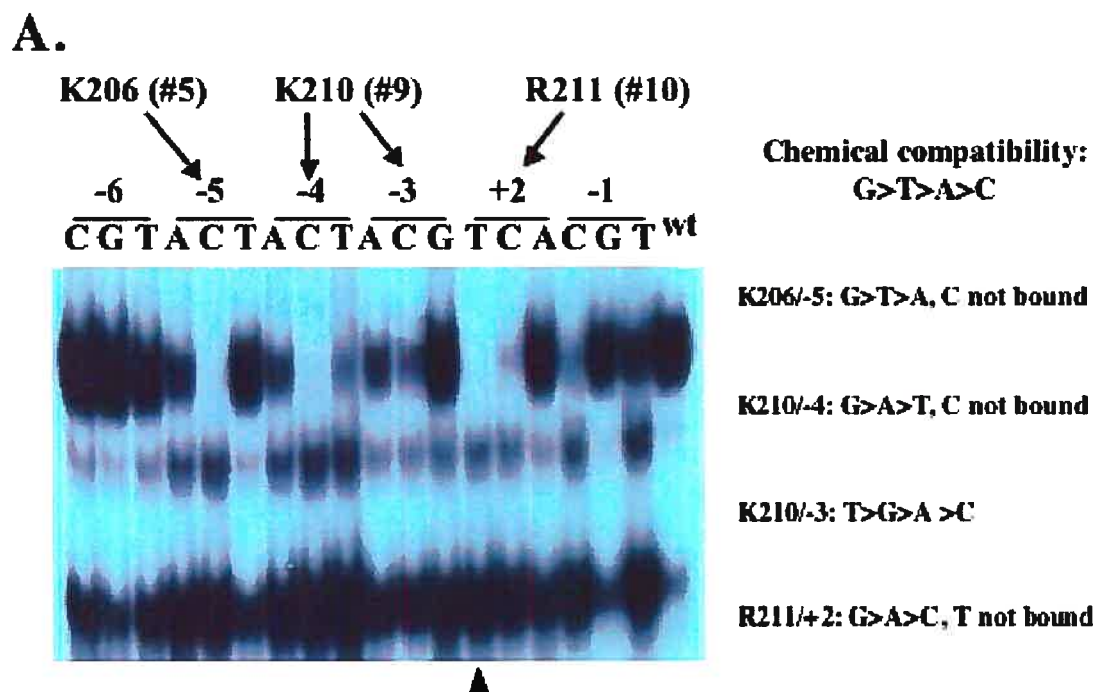


Figure 2.3 (see legend on next page)

**Figure 2.3 (previous page) : Chemical and steric constraints determine the specificity of complex formation with wt ER $\alpha$ .** **A.** Gel shift assays performed with the isolated ER $\alpha$  DBD and a panel of probes containing all possible base replacements introduced at each position of a single motif of the consensus ERE. Contacts with basic amino acids are indicated, and predicted and observed patterns of interactions are described on the right hand side of the figure. **B.** Interaction between R 211 and G+2 as seen in the crystal structure of the ERDBD/ERE (ref, top panel). Note the H bond formed between the amino group of Arg 211 and the N7 atom of G+2. Replacement of G+2 (top panel) by T+2 (middle panel) results in steric conflict between the amino groups of Arg 211 and the methyl group of T+2; energy minimization shows that steric hindrance can be relieved, but Arg 211 cannot interact in the major groove with O4 of T+2. Distances between atoms are represented by a dotted line and a number. The numbers represent the measured distance in Å. **C.** Interaction between K210 and G-4 and T-3 as seen in the crystal structure of the ERDBD/ERE complex (ref, top panel); note the H bonds formed between the amino group of Lys 201 and the N7 atom of G-4 and the O4 atom of T-3; replacement of G-4 (top panel) by T-4 (middle panel) results in steric conflicts between the side chain of Lys 210 and the methyl group of T-4; energy minimization shows that displacement of Lys 210 side chain to accommodate this methyl group results prevents Lys 210 from interacting with charged groups of T-4 and T-3. Distances between atoms are represented by a dotted line and a number. The numbers represent the measured distance in Å.

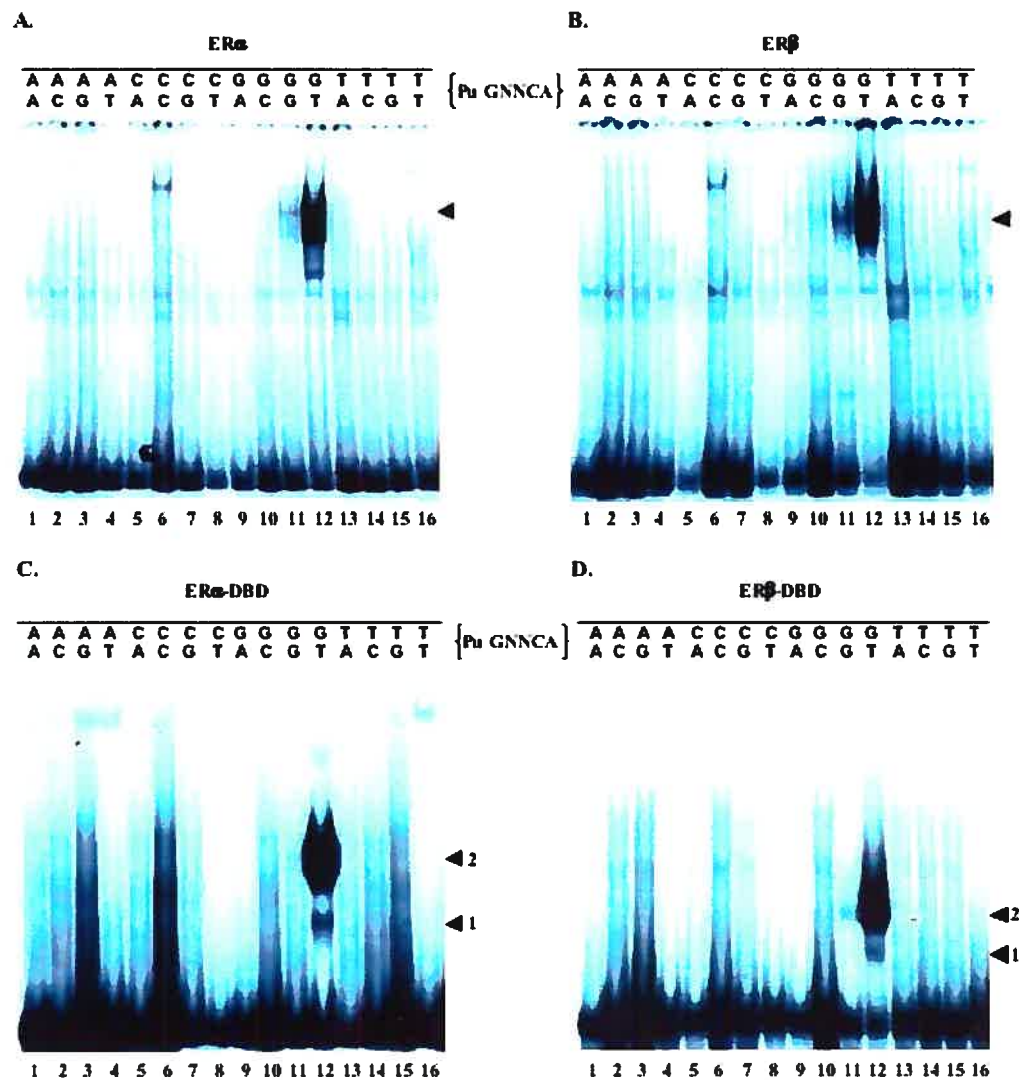
constraints. Replacement in a single motif by T at position -4 drastically reduced binding (figure 2.3A, lane 9) and replacement in both motifs of the ERE completely eliminated complex formation (figure 2.1B-E, lanes 9) and transactivation (data not shown, see also (32)). This contrasts with the capacity of other nuclear receptors such as RAR, RXR and VDR to bind to PuGTTCA motifs (3,32,33), and results from steric constraints within the ER-ERE complex structure. Energy minimization indicates that movement of Lys 210 to avoid the methyl group of T-4 prevents interaction with DNA (figure 2.3C). In addition, Glu 203 is not capable of interacting with the amino group of A at position +4 (not shown). At position -3, G was less favourable than T (compare lanes 12 and 19 in figure 2.3A and figure 2.1B-D), contrary to what is observed at the other positions recognized by basic residues (figure 2.3A). The reason for this observation is that C +3 would create packing

problems with Glu 203, preventing interaction of this amino acid with C+4 (not shown). Thus, our experimental and modeling data suggest that both Glu 203 and Lys 210 contribute to the selectivity of response element recognition with respect to the two central bases of the ERE.

### **2.5.3 Roles of Glu 203 and Lys 210 in the specificity and affinity of ER interaction with palindromic response elements.**

To better analyze the respective roles of Glu 203 and Lys 210 in determining the specificity and affinity of ER interaction with response elements, we further characterized the ER binding specificity with respect to the two central base pairs of the palindrome motifs using a panel of 16 probes representing all combinations of these central bases introduced in both arms of the palindrome. These two base pairs discriminate between elements bound by ER (PuGGTCA motifs) and response elements recognized by GR, AR, PR and MR, and are recognized by different amino acids in ER and GR ((17-19); see also figure 2.2A). Full length estrogen receptors and isolated DBDs bound with high affinity only to the GT (response elements are designated by bases found on the minus strand at position -4/-3) combination found in the wt ERE and, with lower affinity, to the GG element (figure 2.4A-D). Note that complexes formed on element CC with full length receptors (lane 6) were also observed with extracts from cells transfected with the parental expression vector, and that the smears observed in some lanes with bacterially expressed ER DBDs also appeared with extracts from bacteria transformed with the parental expression vector (not shown).

This very high selectivity of ERs for the two central base pairs in their response elements differs from that of the GR or AR, whose DBDs could both recognize all four PuGNACA-containing palindromes ((16) and data not shown). Crystallographic analysis of the GR DBD indicates that contacts are established only with the base at position +3 through a Val residue (position 6 in the DNA binding helix). Glu 203 is replaced by a Gly residue in GR and AR, and the Lys residue corresponding to Lys 210 in ER $\alpha$  does not contribute to GRE recognition. The higher selectivity of ER interaction with response elements could therefore result either from the fact that Glu 203 interacts with both adjacent



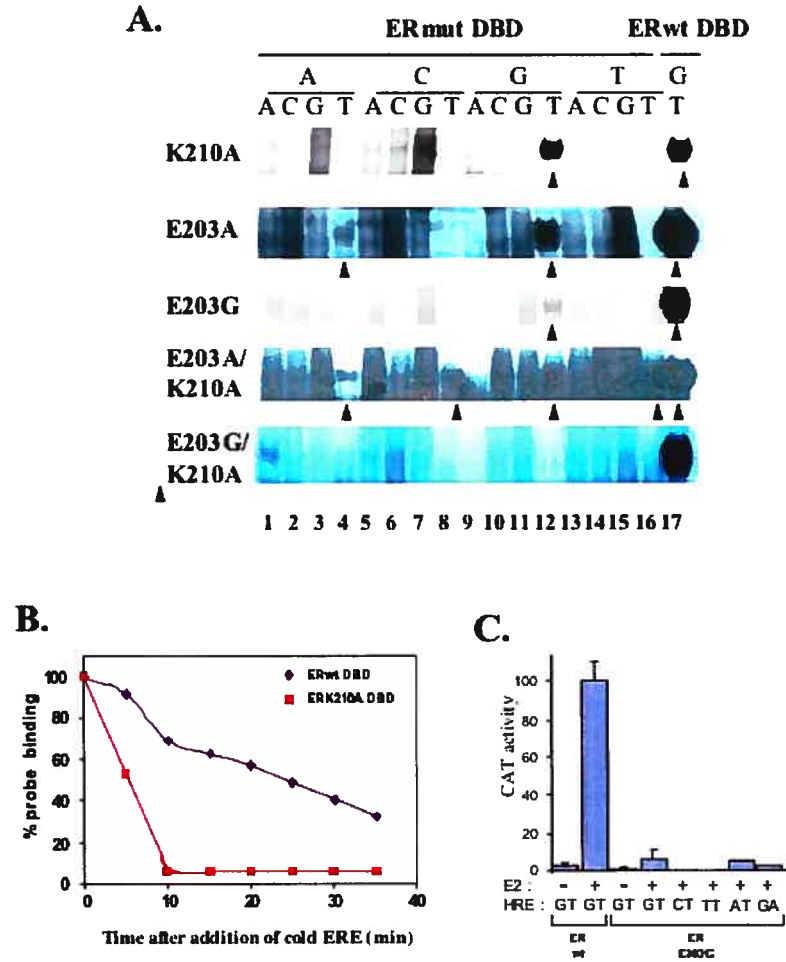
**Figure 2.4 : ERs bind with high selectivity to PuGNNCA palindromes.** A. and B. Extracts from transiently transfected COS-1 cells expressing full length ER $\alpha$  (A) or ER $\beta$  (B) were incubated with  $^{32}$ P-labelled oligonucleotide probes containing PuGNNCA palindromes and complexes were separated on 5% polyacrylamide gels. The position of the dimeric complex is indicated by an arrow. C. and D. Bacterial extracts containing the ER $\alpha$  (C) or ER $\beta$  (D) DNA binding domains were incubated with the same probes and complexes were separated by electrophoresis on 7% polyacrylamide gels. The position of specific complexes containing monomers (1) or dimers (2) of ER DBDs is indicated on the figure.

+4/+3 bases, and/or from the contribution of Lys 210 in recognizing the -4/-3 bases on the opposite strand of DNA. To investigate the respective roles of these two amino acids to the



affinity and/or selectivity of binding, we replaced either Lys 210 or Glu 203 by Ala residues in the ER $\alpha$  DBD and examined complex formation with the panel of 16 probes corresponding to all possible variants at the two central positions of the repeated motifs. The K210A DBD still bound with highest efficiency to the consensus ERE (figure 2.5A, lane 12), although the intensity of the retarded band was lower than with the wt DBD. Accordingly, off rate curves indicate a 4-fold reduction in the amount of time needed to displace 50% of the complex affinity (figure 2.5B). However, the K210A mutation was still fully active in transactivation assays using a reporter gene containing the consensus ERE (figure 2.9B). Interestingly, no other high affinity complexes were observed, suggesting the role of Lys 210 in the selectivity of response element recognition may be mainly redundant with the role played by Glu 203.

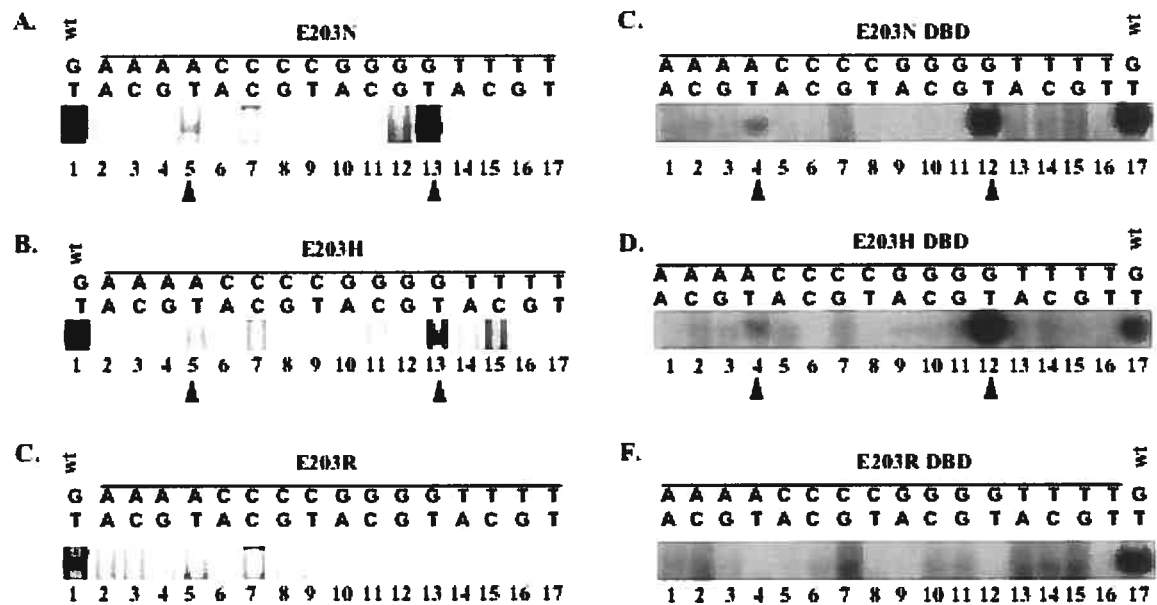
Mutant E203A bound with reduced efficacy to the consensus ERE compared to the wt DBD (figure 2.5A, lane 12), and formed another complex with the PuGATCA probe (AT, figure 2.5A, lane 4), which was not detected with the K210A mutation. Off rate curves indicate that the complex was fully competed by 5 minutes under the experimental conditions, suggesting that this mutation is more destabilizing than the K210A mutation. Mutation E203G also considerably reduced binding efficiency to the consensus ERE, leading to quick off-rates and undetectable transactivation capacity (data not shown and Fig. 2.5C); contrary to mutant E203A, not binding to probe AT was detected. Finally, the double mutant K210A-E203A bound weakly to the cons ERE probe, but also formed detectable complexes with probes AT, CT, and TT (figure 2.5A, lanes 4, 8, 12 and 16). Lack of binding to the CT probe can be explained by charge conflicts between C-4 and Lys 210 and/or G+4 and Glu 203, consistent with a redundant restrictive role of these two residues. Binding of E203A to the AT probe suggests that the main role in restricting binding to this element in the wt receptor is played by Glu 203 rather than Lys 210, and could result from steric conflict between the Glu side chain and T+4. Finally, replacement of both Glu 203 and Lys 210 by Ala residues is apparently necessary to eliminate collision of T-4 with Lys 210 and alleviate the negative role played by the non-neutralized charge of Glu 203 facing A+4 (see above). Thus K210 and E203 appear to have both redundant and complementing roles in restricting binding specificity both through charge and steric conflicts.



**Figure 2.5 : Role of K210 and E203 in the affinity and specificity of receptor-DNA complex formation.** **A.** Gel shift assays performed with mutants of the ER $\alpha$  DBD expressed in *E. coli* and a panel of probes containing all possible base combinations introduced at the two central positions of both PuGGTCA motifs in the consensus ERE. Arrows indicate the main specific complexes between DBD mutants K210A, E203A, and E203A/K210A and some of the 16 probes used in this assay. **B.** Off rate curves obtained with either the wt ER DBD or that of the K210A mutant. **C.** Patterns of transactivation observed after transient transfection of an expression vector for wt ER or the E203A, K210A, or E203AK210A double mutant, reporter vectors containing the GT, CT, TT or AT elements upstream of the thymidine kinase promoter and the CAT gene, and an internal control CMV- $\beta$ Gal reporter vector. Cells were treated with estrogen or vehicle for 48 h and extracts were assayed for  $\beta$ -galactosidase and CAT activity. Standardized CAT activity is shown.

#### **2.5.4 The effects of mutations of Glu 203 on selective DNA binding are not predictable directly from chemical and steric compatibility rules.**

A simple DNA recognition model for steroid receptors has been proposed previously (23), based on both the general rules of chemical compatibility between amino acids and base pairs, and stereochemical constraints due to the position of the DNA binding helix across the major groove as derived from the crystal structures of the ER and GR DBDs (figure 2.2C). Note that this model is based on a one amino acid-one base interaction relationship, and in particular does not take into account recognition of the two central bases by both Glu 203 and Lys 210, considering only the role of Glu 203 with C+4 (figure 2.2B, bold line). Our finding that Glu 203 and Lys 210 cooperate for selective binding to GT palindromes incited us to test the predictions of this model for replacement of Glu 203 by other amino acids. Replacement of Glu by Asn is predicted to switch recognition from C at position +4 to A (TN elements), and replacement by His or Arg should lead to specific recognition of a G at this position (CN elements, figure 2.2C). These mutations were introduced in ER $\alpha$  full length and in the isolated DBD. Similar expression levels were obtained for all full length mutants and the wild-type receptor. When whole cell extracts containing the mutant receptors E203N and E203H were incubated with the wt ERE (GT probe), strong complexes were formed (figure 2.6A and B, lane 13). These complexes can lead to transcriptional activation, as demonstrated by transient cotransfection of CAT reporter vectors containing these response elements cloned upstream of the thymidine kinase promoter with expression vectors for the different mutants (figure 2.8A, ERE-tk-CAT reporter). While binding to the wt ERE was not abolished, the patterns of probe recognition were altered with both mutants. Both mutants E203N and E203H bound to the AT element (figure 2.6A-B, lanes 5 and 3). Note that interaction of these mutants with the AT element were transcriptionally productive, as demonstrated by transient transfection of the full length receptors with a tk reported vector containing AT palindromes (figure 2.8A, AT-tk-CAT reporter vector). Strikingly, these results do not correspond to predictions based on Suzuki's model, as no stable binding was observed with TN motifs for E203N (figure 2.6A, lanes 14-17), or CN motifs with E203H (figure 2.6B, lanes 6-9). Accordingly, the TT and CT elements did not confer estrogen responsiveness to the tk promoter in the presence of these mutants. Note that no specific complexes were observed with the E203R

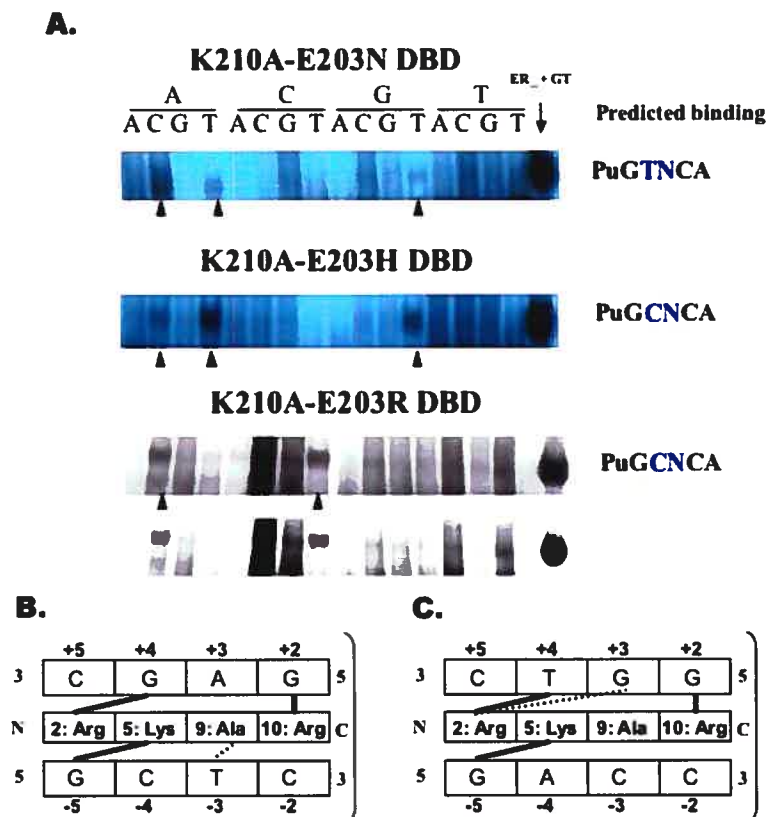


**Figure 2.6 : Mutations of Glu 203 do not generate the expected switch in DNA binding specificity.** A., B. and C. COS-1 cells were transiently transfected with expression vectors for ER $\alpha$  mutant E203N (A) E203H (B) or E203R (C). Whole cell extracts were used in a gel shift assay with the sixteen PuGNNCA probes and complexes were separated by electrophoresis on 5% polyacrylamide gels. The complexes formed are indicated by an arrowhead. For comparison, the complex formed between wt ER $\alpha$  and the consensus ERE (GT) is loaded on each gel (lane 1). D. E. and F. similar assays were performed using bacterial extracts containing mutant ER DBDs carrying the E203R, E203H or E203R mutations.

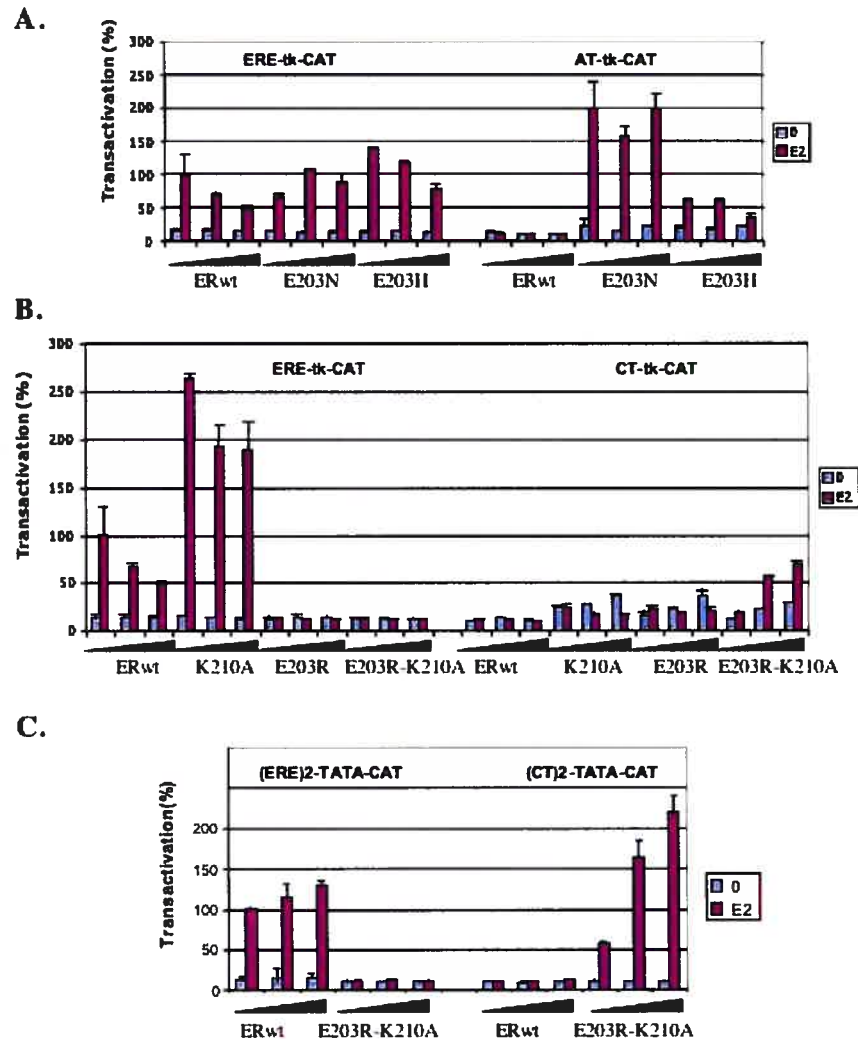
mutant on any of the PuGNNCA probes (figure 2.6C). Lack of binding to the consensus ERE and lack of transactivation on an ERE-tk-CAT reporter (figure 2.8B) suggest that Arg at this position has a destabilizing effect that is stronger than the absence of side chain (Ala or Gly mutations, see figure 2.5). Replacement of Glu 203 by R in the structure of the ER-ERE complex indicates that the Arg side chain is too bulky to fit in the major groove of DNA and that the amino groups exert repulsive effects with the positively charged groups of C+4 and C+5 (figure 2.9C). In addition, neither binding to CN elements or transactivation from reporter vectors containing the CT palindrome (figure 2.8B, CT-

tkCAT reporter vector) were obtained, indicating again that the predicted switch in specificity has not occurred.

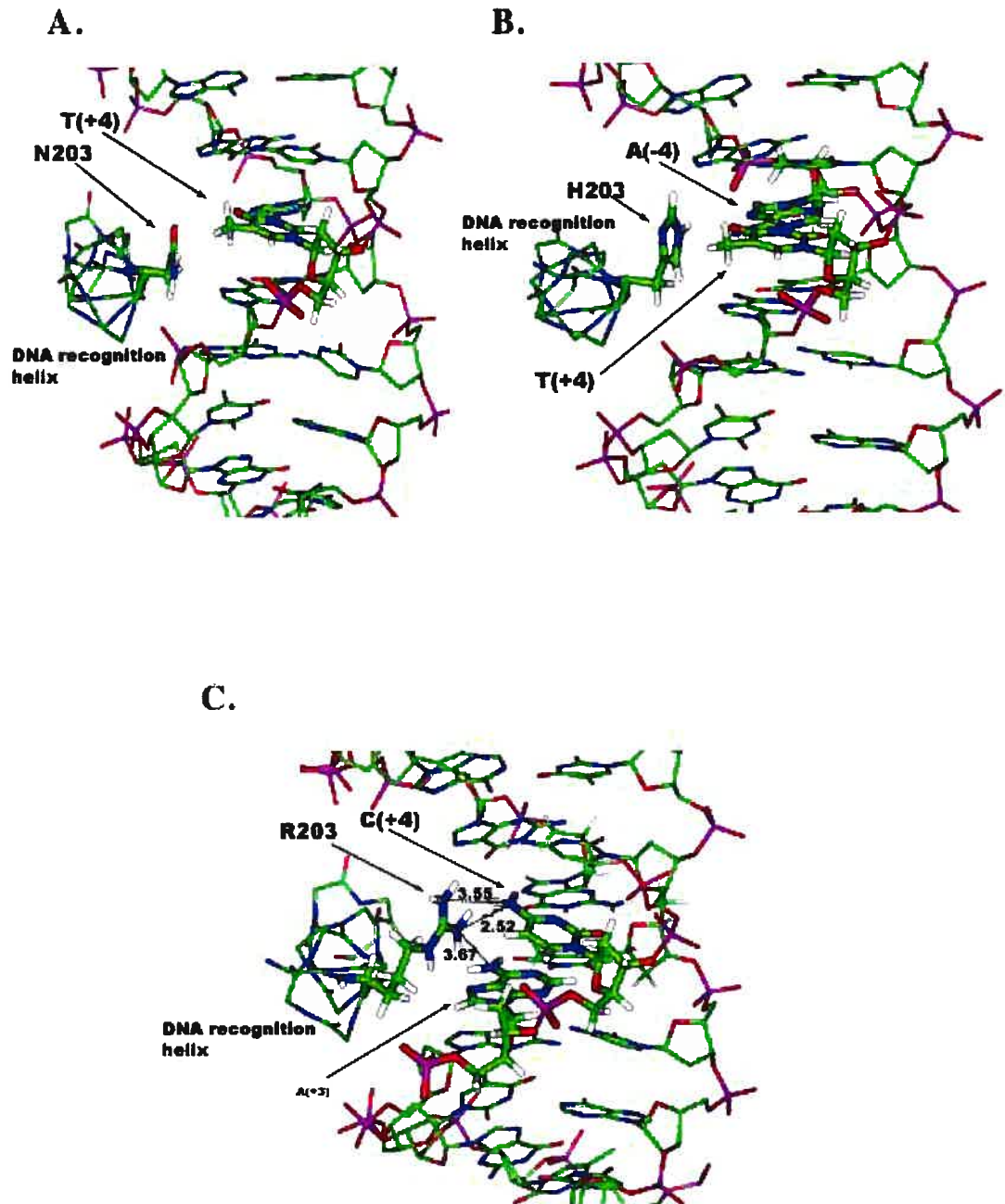
These results indicate that a model based on a one amino acid – one base pair relationship is not an accurate description of the interaction between the estrogen receptor and its target response element, and raise the question of whether Lys 210 plays a role in modulating the DNA binding specificity of receptors carrying mutations at position 203.



**Figure 2.7 : Mutation Lys 210 Ala introduced in Glu 203 mutants reduces binding to the consensus ERE and reveals the switched specificity of mutant E203R. A.** Gel shift assays performed with double mutants of the ER $\alpha$  DBD expressed in *E. coli* and the panel of 16 probes containing all possible base combinations introduced at the two central positions of both PuGGTCA motifs in the consensus ERE. Arrows indicate the main specific complexes formed with DBD mutants E203R-K210A, E203H-K210A, and E203R/K210A **B.** and **C.** Models of interactions in the experimentally observed complexes between E203R-K210A and the CT (**B**) or AC (**C**) elements.



**Figure 2.8 : Altered patterns of reporter vector transactivation by ER mutants. A.** HeLa cells were transiently transfected with an expression vector for the wt ER or the E203N or E203H mutants (add double mutants E230N-K210A, E203H-K210A) and tk-CAT reporter vectors containing the consensus ERE or the AT element, and an internal control CMV- $\beta$ Gal reporter vector. Cells were treated with estrogen or vehicle for 48 h and extracts were assayed for  $\beta$ -galactosidase and CAT activity. Standardized CAT activity is shown. **B.** Same as in **A.** except that expression vectors for the wt ER or the K210A, E203R or E203RK210A mutants were used together with tk-CAT reporter vectors containing the consensus ERE or the AT element, and the internal control CMV- $\beta$ Gal reporter vector. **C.** Same as in **A.** except that expression vectors for the wt ER or the E203RK210A mutants were used together with tk-CAT reporter vectors containing two copies of the consensus ERE or the AT element, and the internal control CMV- $\beta$ Gal reporter vector.



**Figure 2.9 : Modeling of the effect of mutations at position 203 on interaction with palindromic response elements. A.-B.** Complexes formed between the ER mutant E203N (A) or E203H (B) and the AT element after replacement of Glu 203 by Asn or His in the ERDBD/ERE crystal structure and energy minimization. The position of the Asn and His side chains are restrained by the presence of Ser 193. **C.** Steric conflicts and charge repulsion with the amino groups of C+4 and A+3 in the consensus response element resulting from replacement of Glu 203 by Arg. Distances between atoms are represented by a dotted line and a number. The number represents the measured distance in Å.

### **2.5.5 Lys 210 maintains binding to the consensus EREs with mutants at position 203 and prevents binding of mutant E203R to novel response elements.**

Contrary to the total absence of binding observed with the E203R mutant of ER $\alpha$ , mutation of the corresponding amino acid in the androgen receptor (Gly 577) to Arg led to a different pattern of response element binding than with the wt receptor. While wt AR bound all four PuGNACA elements, the G577R mutant bound PuGGACA elements very weakly, but interacted with the three other PuGNACA elements, with a preference for the element containing a G at position +4 (PuGCACA). In the context of the GR DBD, and presumably also the AR DBD, the Lys residue corresponding to Lys 210 (position 9 in the DNA recognition helix) does not bind DNA, but is involved in an interaction with a Glu residue (position 13 in the DNA recognition helix, figure 2.2A). Thus, we investigated whether mutation K210A may facilitate association of the Glu 203 mutants to novel binding sites and/or reduce binding to the consensus ERE. The resulting E203N/K210A and E203H/K210A double mutants had a much reduced binding to this probe, confirming the role of Lys 210 in binding to the consensus element in the absence of a residue at position 203 recognizing C+4. However, these mutants still did not form stable complexes with probes containing bases predicted to interact with the N or H residue at position 203 (TN, CN, respectively). In addition to complexes with AT palindromes already observed with the single mutants, the strongest complex detected was with the AC element. On the other hand, the double mutant E203R-K210A gained binding to the CT element, as could be predicted from chemical compatibility between Arg at position 203 and G+4 (figure 2.7B) and also to the AC element (figure 2.7C). Thus, the specificity of base recognition expected from replacing the Glu 203 residue by Arg was revealed when Lys 210 was mutated to Ala, although an additional unpredicted complex was also formed. As the efficiency of complex formation was about 10-fold lower than that of the wt receptor on the consensus ERE, we examined whether the interaction between the E203R-K210A mutant and the CT probe is transcriptionally productive. Whereas no estradiol-induced transcription could be detected with ERwt, K210A or E203R on a CT-tk-CAT reporter vector, the double mutant E203R-K210A gave rise to a dose-dependent increase in estradiol-stimulated transcription, reaching 70% of the peak estradiol-stimulated transcription with ERwt on ERE-tk-CAT (~2.5 fold induction +/- hormone). The double



mutant like the single mutant E203R was not active with ERE-tk-CAT, confirming the switch in specificity. Similar results were obtained with reporter vectors containing 2 consensus or CT palindromes, except that transactivation with the double mutant was more pronounced.

## 2.6 Discussion

Although palindromic EREs represents the highest affinity element for estrogen receptors, most of the natural response elements characterized so far in the promoters of estrogen target genes contain deviations from the consensus sequence. In view of the large amount of sequence information provided by genome sequencing efforts, it is important to assess the influence of individual base replacements on the affinity of binding by estrogen receptors. Here we have investigated systematically the influence of symmetrical base replacements in both motifs of the consensus palindrome. No major difference in DNA binding selectivity was observed here between ER $\alpha$  and ER $\beta$ . This may not be surprising given that ER $\beta$  contains only two amino acids that differ from ER $\alpha$  in the zinc finger region, and neither of them are in the DNA recognition helix. However, one of the amino acid differences is in the D box, raising the possibility of differences in dimerization properties between the two receptor DBDs. Interestingly, it has been reported that the monomeric SF1 response element was bound by ER $\alpha$  but not by ER $\beta$ . Further experiments are needed to investigate more thoroughly potential differences in the DNA binding patterns of the two receptors on monomeric motifs. The divergence in the amino acid sequences of the two ERs in the region C-terminal to the zinc fingers, which has been involved in recognition of base sequences flanking the PuGGTCA motif and in dimerization in some receptors, may play a role in modulating the affinity of monomeric motif recognition by ER $\beta$ .

Our results indicate that the influence of the diverse replacements in the consensus ERE on binding by ER $\alpha$  and ER $\beta$  varies considerably according to the position in the motif and the nature of the base replacement. Replacement of the Pu residue at position -6 by a Py is tolerated, albeit with a loss in affinity, consistent with the fact that no direct contacts were observed with bases at this position in the crystal structure. A preference for Pu versus Py was also observed at position -1, which is also not directly contacted in the ER $\alpha$  DBD

crystal structure, with a stronger destabilization effect of replacement by Py than at position -6. The reasons for this preference are unclear but may involve effects of base composition on local DNA structure. Replacement of A-1 by G is found in the Cathepsin D, the human lactoferrin and the rat creatine kinase B elements (12), and would be expected to have little destabilizing effect by itself. An A-1 to C replacement is found in the response element of the human mullerian inhibiting substance and in one of the imperfect palindromes in the human TGF $\alpha$  promoter, and would be expected to have a more drastic effect on the affinity of complex formation.

Bases at position -5, -4, -3 and +2 all interact with basic residues, and the preferred bases at these positions are usually G or T, although steric constraints can modulate these preferences, in particular with respect to recognition of T residues. The methyl group of T-4 and T+2 in particular would create steric clashes with interacting amino acids. Finally, ERs display a high degree of specificity for bases at position -4/-3, since GT was the only element bound with high affinity, while GG is tolerated when introduced in only one motif but reduces binding efficacy dramatically when introduced in both motifs. A natural palindromic response element containing a single T-3 to G replacement is found in the promoter of the rabbit uteroglobin gene, which is a strong response element (34). Other elements containing this mutation in combination with other replacements are the human factor XII, rat creatine kinase B and human c-fos elements (12). Replacement of T-3 into C at the same position in the pS2 and the Cathepsin D elements is expected to reduce affinity to a greater extent. It is interesting to note that both these promoters were reported to function poorly in endometrial carcinoma cells, possibly because of the low levels of estrogen receptors in these cells.

Mutation analysis has identified the regions involved in discrimination between an ERE and a GRE (17-19) and in dimerization on the palindromic response element (18,20). Structural analyses of complexes between DNA and the ER $\alpha$  and GR DBDs have identified the molecular interactions that are formed within complexes with their cognate response elements (6,7,35-37). Differences in the specific contacts established by base-interacting amino acids in these two receptors, including those involving conserved amino acids like Lys 210 in ER $\alpha$ , would suggest that the two receptors have a different stringency in their interaction with the two central base pairs and that mutations introduced in residues recognizing these central base pairs may have a different effect on specificity of response

element recognition. Our observation that ERs have a higher specificity for the two central base pairs than GR and AR is therefore in agreement with the structural data. Further, the differential effect of the introduction of Arg at position +2 in the DNA recognition helix in AR, resulting in discrimination towards the base at position +4 (16), and in ER, where this mutation led to an inactive receptor, can also be explained if considering the differential role of Lys at position 5 of the DNA recognition helix (Lys 210 in ER $\alpha$ ) in the two receptors. The corresponding Lys in GR/AR does not interact with DNA but rather with glutamic acid at position 13 of the DNA recognition helix, possibly due to the different positioning of the DNA helix observed in the crystal structures of the ER and GR DBDs. Although mutation of Lys 210 to alanine had only a minor effect on the stability of the interaction with the wild type receptor and did not alter the specificity of DNA recognition, differences in the binding patterns of the E203A and E203A-K210A mutants revealed the complementary role of these amino acids in maintaining binding specificity to the wild type response element through charge and steric constraints.

A simple model for prediction of specific amino acid-base interactions in the context of the steroid receptor DNA binding helix has been proposed (23). In spite of the fact that this model incorporates both chemical rules governing amino acid-base interactions and stereochemical constraints resulting from the position of the DNA recognition helix in the major groove of response element, we have found that none of the elements predicted to interact with mutants E203N, E203H or E203R (TN, CN, and CN, respectively) were bound. Both E203N and E203H still interacted with high affinity with the wild type response elements (GT). On the other hand, interactions with the AT element, which was not predicted by the proposed model, were observed in gel shift and transactivation assays with both E203N and E203H mutants. The fact that the E203A mutant also recognized the AT element suggests that Glu 203 actively represses binding to elements containing T at position +4, due to negative interactions between the carbonyl group and the methyl group of thymine. Our modeling indicates that the methyl group of thymine can be accommodated by side chain rearrangement when Glu 203 is replaced by Asn or His, even though these residues do not make hydrogen bonds with T+4. The Arg side chain is also too bulky for interaction with T+4 but could interact with the phosphate at position +3, explaining the observed binding of the E203R-K210A to the AA element.

The role of Lys 210 in preventing a specificity switch by mutation at position 203 is demonstrated by the absence of specific interactions observed with the E203R mutant, whereas the double mutant E203R-K210A could bind the CT element, as predicted by base compatibility between Arg and G+4. The presence of Lys 210 indeed results in a conflict for binding to bases with opposite charges. As noted above, a Lys exists in AR at the corresponding position but does not contact DNA, explaining the capacity of the AR mutant with an Arg at the position equivalent to 203 to interact with G+4. Another difference between the two receptors is the contribution of Val at position 6 in the DNA recognition helix (figure 2.2A) in stabilizing the complex through recognition of T+3. In the context of the ER, in which Val is replaced by Ala, both the lack of a stabilizing interaction with the base at position +3 and the conflicting preferences of Arg and Lys for the -4/-3 and +4/+3 bases explain the total lack of binding on any PuGNNCA element of the single E203R mutant. The double mutant E203R-K210A exhibited a total switch in DNA binding specificity, with lack of binding to the consensus probe due to charge and steric conflict of Arg with C+4 (figure 2.9C) and transcriptionally productive interactions with the CT palindrome.

Mutation K210A also did not reveal expected interactions between Asn 203 and elements containing an A+4 (TN elements). Binding to TT elements was only observed with the E203A-K210A mutant. This is surprising since several nuclear receptors capable of recognizing PuGTTCA elements share both these amino acids with ER. In addition, a GR DBD containing the three P box amino acids of ER was capable of interacting with the TT probe more efficiently than with the consensus ERE; binding to TT was also observed in mutants of this chimeric DBD where the first amino acid of the ER P box (Glu) was replaced by Arg, Asn, Ala, and Gly (38). Of interest, the Lys residue corresponding to amino acid 210 does not contact DNA in GR, due to differences in the positioning of the GR DNA binding helix (39), and those in the RAR-RXR heterodimer make only a weak contact with G-4 in a direct repeat element composed of two PuGGTCA motifs separated by 1 bp (40).

All together, our results indicate that simple chemical and stereochemical rules cannot predict accurately the changes in the selectivity of estrogen receptor/DNA interactions induced by specific mutations, at least when considering the two central base pairs. Indeed, predicted interactions were not observed while unpredicted complexes were

formed. An accurate model for ER/DNA interactions needs to incorporate both recognition of adjacent bases by one single amino acid, and recognition of base pairs by two amino acids interacting on opposite strands of DNA. The combined effects of Glu 203 and Lys 210 in interacting with the same base pair is apparent both at the level of charge and steric constraints, resulting in the tighter DNA binding specificity for the two central base pairs observed for ER versus other steroid receptors. Further, alterations in other residues of the DNA binding helix and/or differences in the mode of DBD dimerization all have the potential to modulate the relative positioning of base-interacting amino acids with respect to the DNA and thus to alter the patterns of response element recognition obtained with similar mutations in different receptors.

## 2.7 Acknowledgements

This work was supported by an NSERC grant to SM. JD is recipient of an FRSQ fellowship. SC is a CIHR scientist and SM is an FRSQ Chercheur-Boursier Senior and holds the CIBC Breast Cancer Research Chair at Université de Montréal.

## 2.8 References

1. Green, S. and Chambon, P. (1988) *Trends Genet*, **4**, 309-314.
2. Beato, M. and Klug, J. (2000) *Hum Reprod Update*, **6**, 225-236.
3. Glass, C.K. (1994) *Endocrine Rev*, **15**, 391-407.
4. Rastinejad, F., Perlman, T., Evans, R.M. and Sigler, P.B. (1995) *Nature*, **375**, 203-211.
5. Khorasanizadeh, S. and Rastinejad, F. (2001) *Trends Biochem Sci*, **26**, 384-390.
6. Luisi, B.F., Xu, W.X., Otwinowski, Z., Freedman, L.P., Yamamoto, K.R. and Sigler, P.B. (1991) *Nature*, **352**, 497-505.
7. Schwabe, J.W., Chapman, L., Finch, J.T. and Rhodes, D. (1993) *Cell*, **75**, 567-578.
8. Klein-Hitpass, L., Schorpp, M., Wagner, U. and Ryffel, G.U. (1986) *Cell*, **46**, 1053-1061.
9. Driscoll, M.D., Sathya, G., Muyan, M., Klinge, C.M., Hilf, R. and Bambara, R.A. (1998) *J Biol Chem*, **273**, 29321-29330.
10. Klinge, C.M. (2001) *Nucleic Acids Res*, **29**, 2905-2919.
11. Walker, P., Germond, J.-E., Brown-Luedi, M., Givel, F. and Wahli, W. (1984) *Nucleic Acids Res*, **12**, 8611-8626.
12. Sanchez, R., Nguyen, D., Rocha, W., White, J.H. and Mader, S. (2002) *Bioessays*, **24**, 244-254.
13. Bourdeau, V., Deschenes, J., Metivier, R., Nagai, Y., Nguyen, D., Bretschneider, N., Gannon, F., White, J.H. and Mader, S. (2004) *Mol Endocrinol*, **18**, 1411-1427.

14. Klock, G., Strähle, U. and Schütz, G. (1987) *Nature*, **329**, 734-736.
15. Nordeen, S.K., Suh, B.J., Kühnel, B. and Hutchinson, C.A. (1990) *Mol Endocrinol*, **4**, 1866-1873.
16. Nguyen, D., Steinberg, S.V., Rouault, E., Chagnon, S., Gottlieb, B., Pinsky, L., Trifiro, M. and Mader, S. (2001) *Mol Endocrinol*, **15**, 1790-1802.
17. Mader, S., Kumar, V., de Verneuil, H. and Chambon, P. (1989) *Nature*, **338**, 271-274.
18. Umesono, K. and Evans, R.M. (1989) *Cell*, **57**, 1139-1146.
19. Danielsen, M., Hinck, L. and Ringold, G.M. (1989) *Cell*, **57**, 1131-1138.
20. Mader, S., Chambon, P. and White, J.H. (1993) *Nucleic Acids Res*, **21**, 1125-1132.
21. Gewirth, D.T. and Sigler, P.B. (1995) *Nat Struct Biol*, **2**, 386-394.
22. Schwabe, J.W., Chapman, L. and Rhodes, D. (1995) *Structure*, **3**, 201-213.
23. Suzuki, M. and Yagi, N. (1994) *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 12356-12361.
24. Segal, D.J., Beerli, R.R., Blancafort, P., Dreier, B., Effertz, K., Huber, A., Kokschi, B., Lund, C.V., Magnenat, L., Valente, D. *et al.* (2003) *Biochemistry*, **42**, 2137-2148.
25. Sera, T. and Uranga, C. (2002) *Biochemistry*, **41**, 7074-7081.
26. Blancafort, P., Segal, D.J. and Barbas, C.F., 3rd. (2004) *Mol Pharmacol*, **66**, 1361-1371.
27. Tora, L., Mullick, A., Metzger, D., Ponglikitmongkol, M., Park, I. and Chambon, P. (1989) *Embo J*, **8**, 1981-1986.
28. Klein-Hitpass, L., Ryffel, G.U., Heitlinger, E. and Cato, A.C. (1988) *Nucleic Acids Res*, **16**, 647-663.
29. Barsalou, A., Gao, W., Anghel, S., Carriere, J. and Mader, S. (1998) *J. Biol. Chem.*, **273**, 17138-17146.
30. Pearlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.S., Cheatham, T.E.I., Ferguson, D.M., Seibel, G.L., Singh, U.C., Weiner, P.K. and Kollman, P.A. (1995).
31. Brzozowski, A.M., Pike, A.C., Dauter, Z., Hubbard, R.E., Bonn, T., Engstrom, O., Ohman, L., Greene, G.L., Gustafsson, J.A. and Carlquist, M. (1997) *Nature*, **389**, 753-758.
32. Mader, S., Leroy, P., Chen, J.Y. and Chambon, P. (1993) *J Biol Chem*, **268**, 591-600.
33. Mader, S., Chen, J.Y., Chen, Z., White, J., Chambon, P. and Gronemeyer, H. (1993) *Embo J*, **12**, 5029-5041.
34. Slater, E.P., Redeuilh, G., Theis, K., Suske, G. and Beato, M. (1990) *Mol. Endocrinol.*, **4**, 604-610.
35. Schwabe, J.W., Neuhaus, D. and Rhodes, D. (1990) *Nature*, **348**, 458-461.
36. Baumann, H., Paulsen, K., Kovacs, H., Berglund, H., Wright, A.P., Gustafsson, J.A. and Hard, T. (1993) *Biochemistry*, **32**, 13463-13471.
37. van Tilborg, M.A., Bonvin, A.M., Hard, K., Davis, A.L., Maler, B., Boelens, R., Yamamoto, K.R. and Kaptein, R. (1995) *J Mol Biol*, **247**, 689-700.
38. Zilliacus, J., Wright, A.P., Carlstedt-Duke, J., Nilsson, L. and Gustafsson, J.A. (1995) *Proteins*, **21**, 57-67.
39. Luisi, B.F., Schwabe, J.W. and Freedman, L.P. (1994) *Vitam Horm*, **49**, 1-47.
40. Rastinejad, F., Wagner, T., Zhao, Q. and Khorasanizadeh, S. (2000) *Embo J*, **19**, 1045-1054.

**CHAPITRE 3 : MATÉRIEL ET MÉTHODES; RÉSULTATS, 2<sup>ÈME</sup> PARTIE**

### 3.1 Matériel et méthodes

Les structures utilisées pour l'analyse structurale sont disponibles en format PDB sur le site Web du Protein Data Bank (<http://www.rcsb.org/pdb/>). Pour la structure du ERDBD en interaction avec le ERE [21], le code d'accèsion est le 1HCQ (résolution : 2,4Å) et pour la structure du GRDBD en interaction avec le GRE [13], le code d'accèsion est le 1GLU (résolution : 2,9Å). Dans le cas du GRDBD, une structure dont les coordonnées ont été raffinées à une résolution de 2,5Å par W. Xu et B. Luisi au laboratoire de Paul Sigler à l'Université de Yale a été utilisée. Les coordonnées de cette structure nous ont été aimablement fournies par Daniel Gewirth, de l'Université de Duke.

L'analyse visuelle des structures a été faite en utilisant un ordinateur O2 de Silicon Graphics. Les superpositions des structures et les calculs de RMSD ont été faits en utilisant les outils fournis dans le logiciel InsightII (Accelrys; San Diego, CA). L'existence de ponts H a été déterminée en mesurant la distance entre les atomes donneurs et accepteurs de ponts H avec les outils fournis dans le logiciel InsightII. Un pont H a été considéré comme existant si la distance entre un atome donneur et un atome accepteur se situait entre 2,60Å et 3,30Å. La modélisation moléculaire de certains complexes a été effectuée en remplaçant les acides aminés d'intérêt par une alanine ou une glycine selon le cas. La structure de l'ADN a été fixée pour ne pas tenir compte des variations possibles dans la conformation de l'ADN lors de la minimisation d'énergie. Cette dernière a été effectuée en utilisant le champs de force AMBER [30] jusqu'à ce qu'un minimum énergétique ait été atteint. La modélisation moléculaire et la minimisation d'énergie ont été effectuées en utilisant les modules Biopolymer et Discover du logiciel InsightII. Toutes les figures présentées dans ce chapitre ont été générées à l'aide du logiciel InsightII à partir des modèles réalisés pendant cette étude.



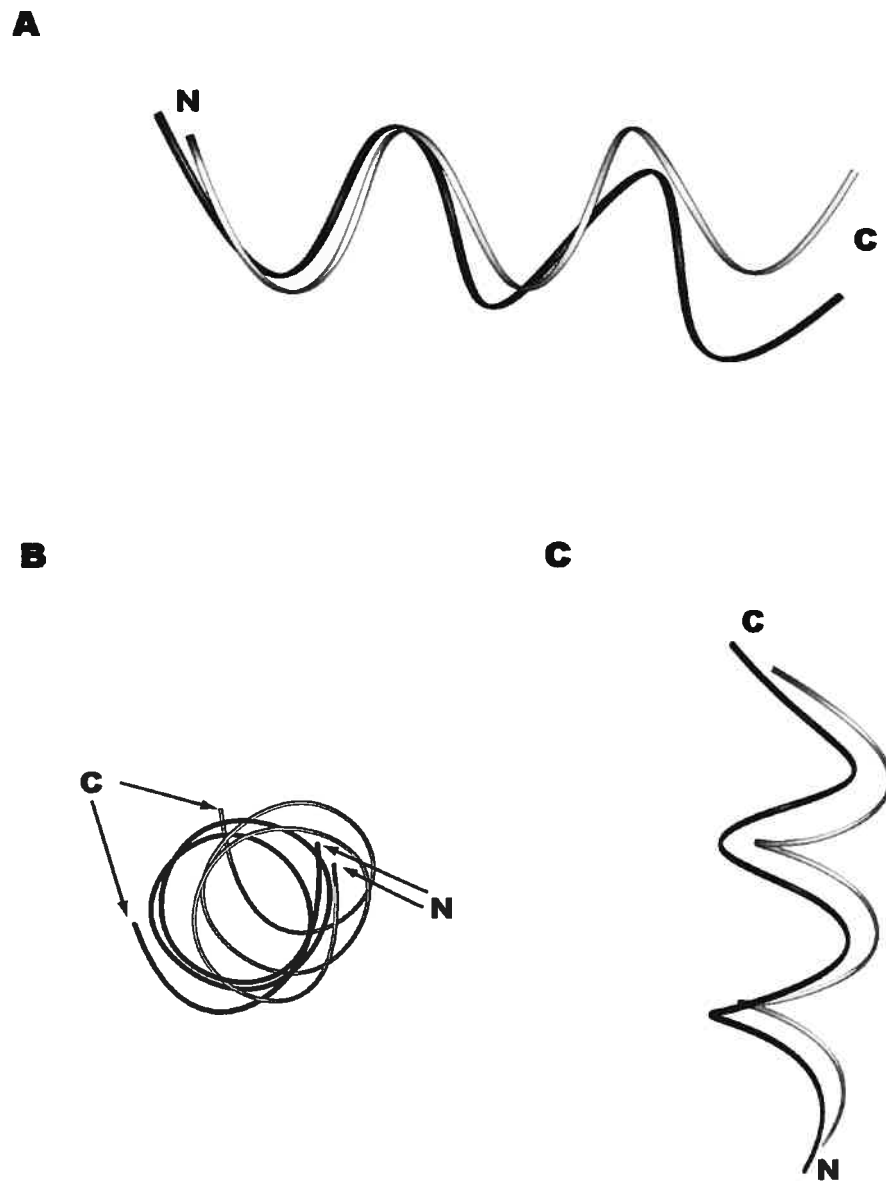
## 3.2 Résultats

### 3.2.1 Différences dans le positionnement de l'hélice de reconnaissance de l'ADN (H1) du GRDBD et du ERDBD dans le sillon majeur de l'ADN

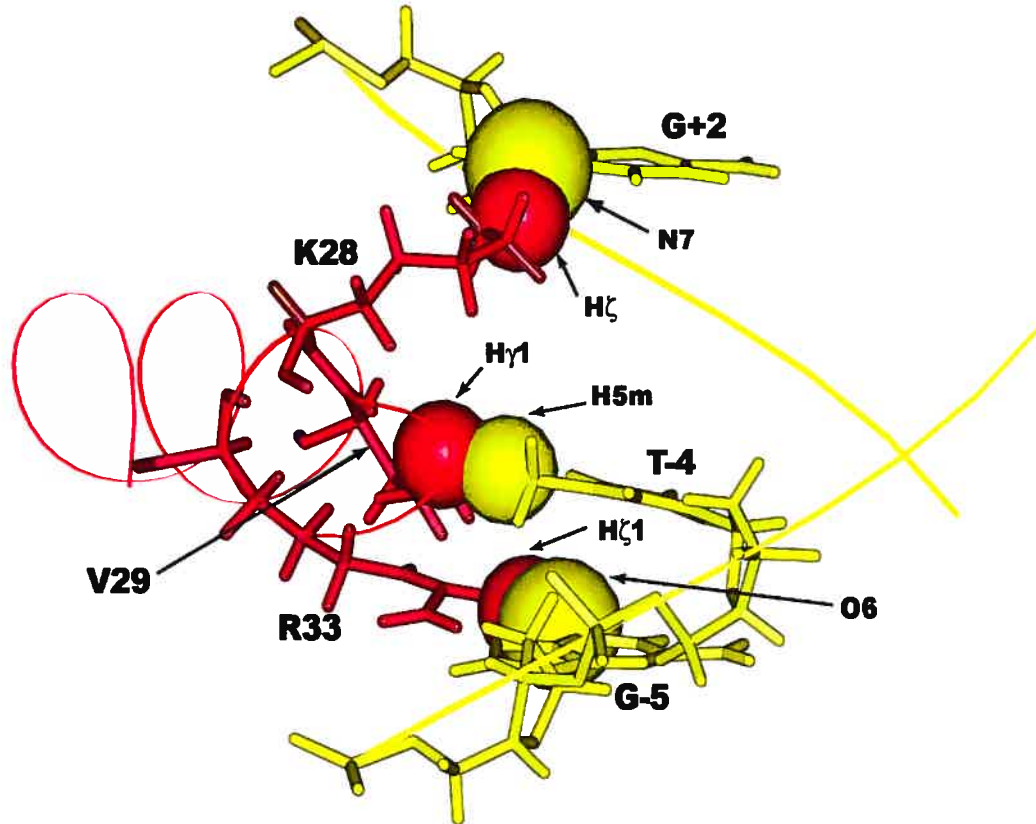
Il a déjà été observé que H1 du GRDBD et du ERDBD sont positionnées différemment dans le sillon majeur de l'ADN dans leur structure respective, sans pour autant que cet aspect soit développé plus en détails [31]. Pour confirmer cette observation et pour en analyser les conséquences, les structures du ERDBD et du GRDBD, en interaction avec leur HRE, ont été superposées. En fait, pour analyser les différences entre les deux structures de la partie protéique des complexes récepteur-ADN, il aurait été préférable de superposer la partie ADN de ces complexes. Par contre, les structures de l'ADN dans ces complexes ne sont pas exactement identiques à cause des différences locales dans la conformation de l'ADN. Pour ignorer les différences de conformation dans la structure de l'ADN, les liens glycosidiques des nucléotides formant les deux paires de bases centrales ont été superposés avec un 'root mean square deviation' (RMSD) de 0,48Å. Ceci permettait de distinguer les différences dans les structures au niveau de l'interface d'interaction protéine-ADN. Une analyse visuelle de ces complexes superposés a permis de constater que les DBDs des récepteurs n'étaient pas parfaitement superposés, en particulier dans la région de H1. La figure 3.1A montre que H1 du GRDBD et du ERDBD ne sont pas exactement parallèles l'une à l'autre, en particulier dans la région C-terminale. Les figures 3.1B et 3.1C montrent que H1 du ERDBD (ruban gris) est plus près de l'ADN que H1 du GRDBD (ruban noir). Ces résultats ont donc démontré que le ERDBD et le GRDBD présentaient leur H1 dans le sillon majeur de leur HRE de manière différente.

### 3.2.2 Conséquences d'un changement de position de H1 pour la liaison à l'ADN

Pour analyser l'impact de cette différence de positionnement sur l'interaction avec l'ADN, l'H1 d'un récepteur a été déplacée dans la position qu'occupe l'H1 du deuxième récepteur, et vice versa. Dans le cas du GRDBD, la figure 3.2 montre que le déplacement de H1 dans la position occupée par H1 du ERDBD a causé plusieurs collisions entre les



**Figure 3.1 : Différences dans la position de H1 du GRDBD et du ERDBD alors que les paires de bases centrales de leur HRE respectif sont superposées. A : Vue de la face hydrophile (qui interagit avec l'ADN) des H1. B : Vue au travers des H1 en partant de l'extrémité N-terminale vers l'extrémité C-terminale. Les molécules d'ADN se situent à la droite des hélices  $\alpha$  (non montrées). Il s'agit de la même orientation qu'à la figure 1.7. C : Vue du dessus des H1. Les molécules d'ADN se situent à la droite des hélices  $\alpha$  (non montrées). Les H1 sont représentées par un ruban noir pour le GRDBD et par un ruban gris pour le ERDBD. N, extrémité N-terminale; C, extrémité C-terminale.**



**Figure 3.2 : Collisions entre les atomes des acides aminés du GRDBD et les atomes des bases du GRE causées par le déplacement de H1 du GRDBD dans la position occupée par H1 du ERDBD. Le GRDBD est coloré en rouge et le GRE est coloré en jaune. Le squelette de H1 est représenté comme un ruban ainsi qu'une partie du squelette sucre-phosphate du GRE. Seulement les acides aminés et les bases ayant un atome impliqué dans une collision sont montrés sous forme de bâtonnets. Les atomes impliqués dans une collision sont représentés sous forme de sphères. Le volume de chaque sphère est proportionnel au rayon de van der Waals de l'atome.**

acides aminés et les bases qui interagissaient ensemble au préalable. Dans le cas du ERDBD, le tableau III montre que le déplacement de H1 dans la position occupée par H1 du GRDBD a empêché la formation de certains ponts H dans l'interface d'interaction protéine-ADN. Ces résultats ont donc démontré qu'un récepteur ne pouvait pas interagir avec son HRE si H1 n'est pas positionnée correctement dans le sillon majeur de l'ADN.

**Tableau III : Ponts H existants entre les acides aminés du ERDBD interagissant directement avec des bases du ERE en fonction de la position de H1 du ERDBD.** Les distances sont mesurées en Å entre les atomes donneur et accepteur de ponts H. Les acides aminés et les bases impliquées dans les ponts H sont indiqués entre parenthèses.

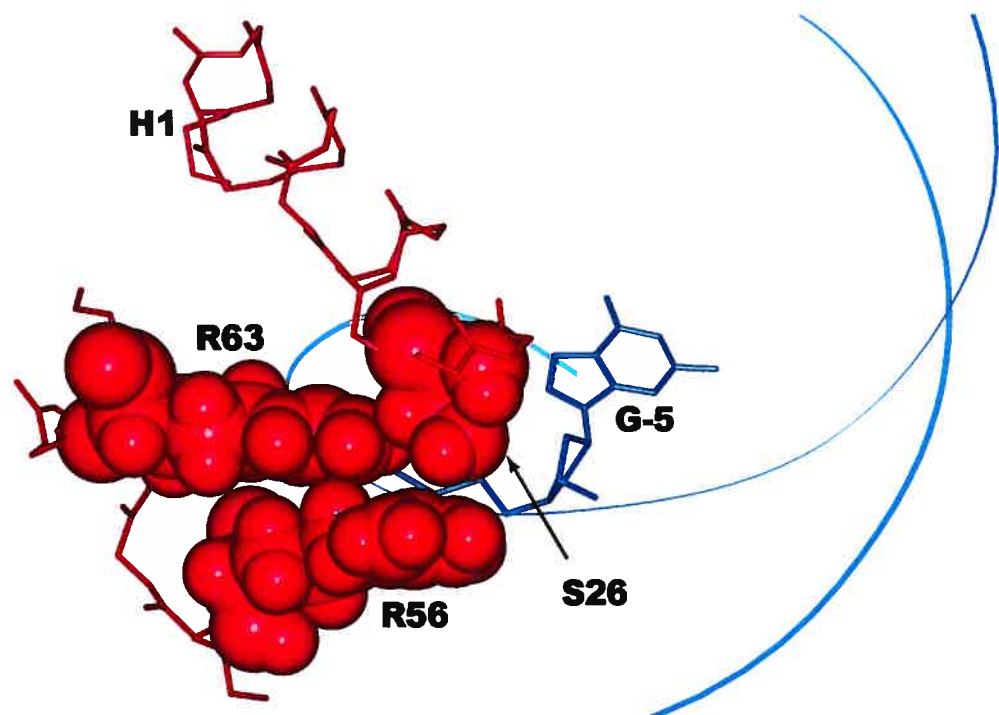
Ponts H	Dans sa position originale	Dans la position de H1 du GRDBD
(Glu25) Oε1 --- N4 (C-3)	3,00	3,74
(Lys28) Nζ --- O6 (G+2)	2,71	4,00
(Lys28) Nζ --- N7 (G+2)	3,15	2,97
(Lys32) Nζ --- N7 (G+3)	3,28	4,32
(Lys32) Nζ --- O4 (T+4)	3,30	4,83
(Arg33) Nη1 --- N7 (G-5)	3,20	3,25

### 3.2.3 Acides aminés responsables du positionnement de H1 dans le sillon majeur de l'ADN

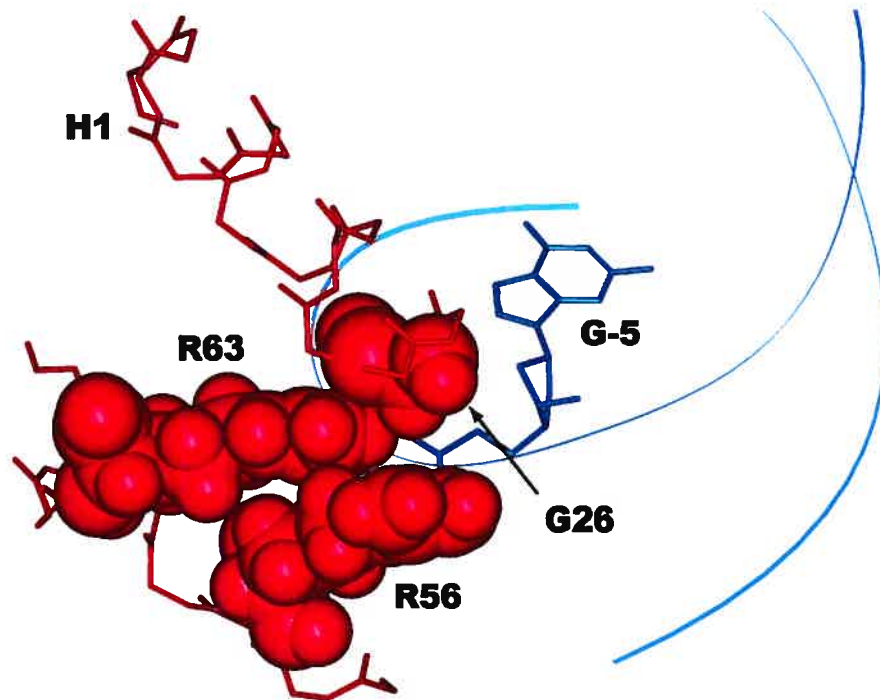
À l'extérieur de la boîte P, il y a trois acides aminés basiques (Lys28, Lys32 et Arg33) qui forment des ponts H avec des bases du HRE et ce, dans les deux structures à l'étude. La seule exception est Lys32 dans le GRDBD qui ne contacte pas directement l'ADN mais qui forme un pont salin avec Gln36. Malgré le fait qu'ils interagissent avec des bases du HRE, ces trois acides aminés ne sont pas considérés comme des déterminants de la sélectivité de la liaison à l'ADN [17 – 19]. Des expériences de minimisation d'énergie pour le GRDBD et le ERDBD en interaction avec leur HRE ont montré que le remplacement des trois acides aminés basiques par des alanines n'affectait pas le positionnement global de H1 dans le sillon majeur de l'ADN (résultats non montrés). Par contre, dans le cas du ERDBD, l'extrémité C-terminale de H1 s'approchait de l'ADN (résultat non montré). Ces résultats ont démontré que Lys28, Lys32 et Arg33 n'étaient pas responsables du positionnement de H1 dans le sillon majeur de l'ADN, à l'exception de Lys32 qui semblait dicter le positionnement de l'extrémité C-terminale de H1 dans le ERDBD seulement.

Comme il a été mentionné à la section 1.5, un aspect intéressant de la boîte P est la présence d'une glycine soit en première ou en deuxième position de la boîte P. De plus, les

acides aminés en deuxième position de la boîte P (Ser26 pour le GR et Gly26 pour le ER) ne forment pas de contact direct avec l'ADN même s'ils sont d'importants déterminants de la sélectivité de la liaison à l'ADN [17 – 19]. Ser26 interagit avec Arg56 dans la structure du complexe GRDBD-GRE [13]. La région de l'acide aminé Arg56 a donc été analysée et comparée avec la même région dans le complexe ERDBD-ERE. L'analyse visuelle de ces régions a montré que deux arginines (Arg56 et Arg63) situées dans le deuxième doigt de zinc contactaient le squelette sucre-phosphate du HRE de manière semblable dans les deux complexes. Ces deux arginines sont conservées à 100% parmi la famille des récepteurs nucléaires. La figure 3.3 montre comment Ser26 est empaquetée sur la chaîne latérale planaire de Arg56. Dans le ERDBD, c'est Gly26, même si elle ne possède pas de chaîne latérale, qui est empaquetée sur la chaîne latérale de Arg56, comme montré à la figure 3.4.

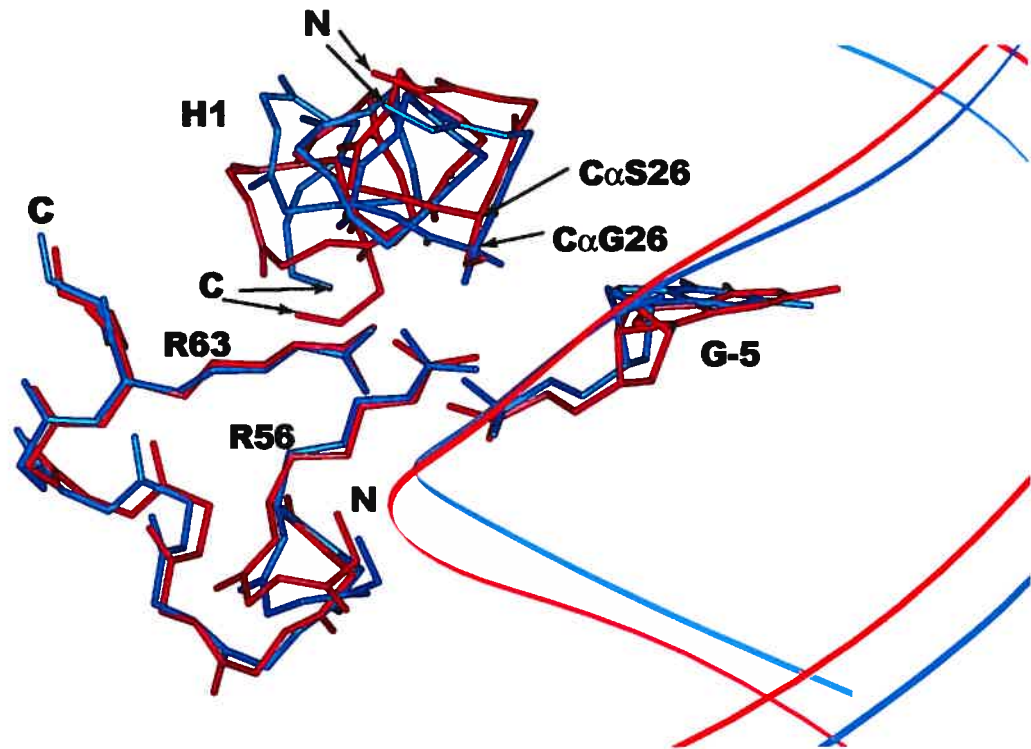


**Figure 3.3 : Entassement de Ser26 sur Arg56 et Arg63 dans le GRDBD.** Le GRDBD est coloré en rouge et le GRE en bleu. Le squelette sucre-phosphate du GRE est représenté comme un ruban. Les squelettes de H1 et du fragment comprenant les acides aminés 54 à 64 du GRDBD sont représentés en bâtonnets, ainsi que le nucléotide G-5. Les atomes des acides aminés Ser26, Arg56 et Arg63 sont représentés en sphères dont le volume est proportionnel au rayon de van der Waals de l'atome.



**Figure 3.4 : Entassement de Gly26 sur Arg56 et Arg63 dans le ERDBD.** Le ERDBD est coloré en rouge et le ERE en bleu. Le squelette sucre-phosphate du ERE est représenté comme un ruban. Les squelettes de H1 et du fragment comprenant les acides aminés 54 à 64 du ERDBD sont représentés en bâtonnets, ainsi que le nucléotide G-5. Les atomes des acides aminés Gly26, Arg56 et Arg63 sont représentés en sphères dont le volume est proportionnel au rayon de van der Waals de l'atome.

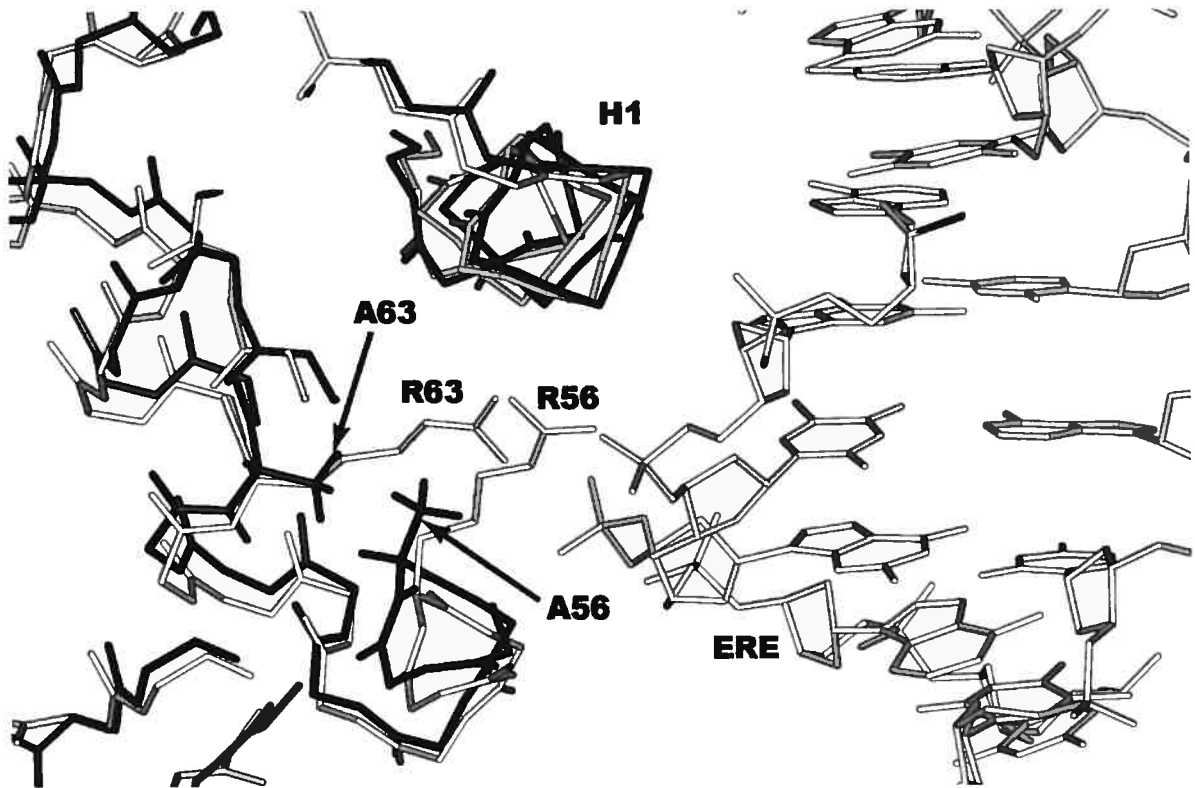
Il a aussi été observé qu'il n'y avait presque plus d'espace libre entre le carbone  $\alpha$  de Gly26 et les chaînes latérales des arginines 56 et 63 (figure 3.4). Donc, si Gly26 et Ser26 étaient entassées sur les mêmes arginines, l'hélice  $\alpha$  du ERDBD devait être décalée par rapport à l'hélice  $\alpha$  du GRDBD pour compenser l'absence de chaîne latérale de Gly26. Pour vérifier ceci, les chaînes latérales de Arg56 et Arg63, ainsi que le groupement phosphate de G-5 du complexe GRDBD-GRE ont été superposés sur les atomes correspondants du complexe ERDBD-ERE avec un RMSD de 0,28Å. La figure 3.5 montre qu'il y avait une différence dans la position des H1 à l'extrémité N-terminale car les hélices n'étaient pas superposées l'une sur l'autre. Des expériences de minimisation d'énergie pour le GRDBD et le ERDBD en interaction avec leur HRE ont montré que le remplacement de Arg56 et Arg63 par des alanines affectait le positionnement global de H1 dans le sillon majeur de l'ADN (figure



**Figure 3.5 : Différences de positionnement des H1 du GRDBD et du ERDBD lorsque Arg56, Arg63 et le groupement phosphate de G-5 sont superposés. Le GRDBD et le GRE sont colorés en rouge alors que le ERDBD et le ERE sont colorés en bleu. Les squelettes sucre-phosphate du GRE et du ERE sont représentés en rubans. Les extrémités N et C terminales sont identifiées par la lettre N et C respectivement. Pour aérer la figure, les squelettes des chaînes polypeptidiques de H1 et du fragment comprenant les acides aminés 54 à 64 sont représentés sous forme de bâtonnets. Le nucléotide G-5 du GRE et du ERE est aussi représenté sous forme de bâtonnets.**

3.6). Pour analyser plus en détails l'interaction entre l'acide aminé en position 26 et les arginines en position 56 et 63, Gly26 a été remplacé par une alanine dans le ERDBD. Contrairement à ce qui était attendu, aucun changement significatif n'a été observé dans le positionnement de H1 (résultat non montré). Ceci pourrait s'expliquer par le fait que le positionnement de H1 est le résultat d'une synergie entre plusieurs acides aminés et révèle la limite de la méthode expérimentale utilisée pour notre recherche (voir section 4.3). Ces résultats ont démontré que l'identité de l'acide aminé en position 26 était un des facteurs déterminant le positionnement de l'extrémité N-terminale de H1 dans le sillon majeur de l'ADN par l'intermédiaire de ses contacts avec les acides aminés Arg56 et Arg63 qui, par





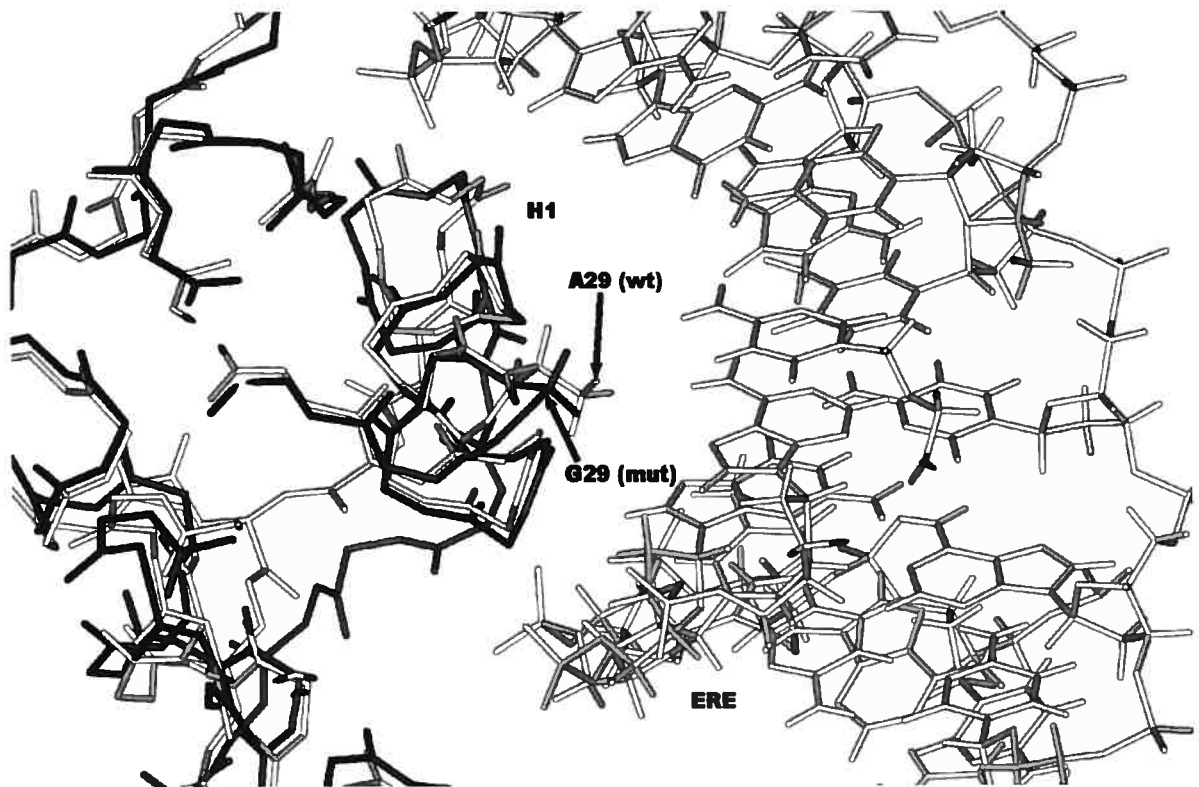
**Figure 3.6 : Déplacement de H1 du ERDBD après le remplacement de Arg56 et Arg63 par des alanines.** Le ERDBD de type sauvage et le ERE sont représentés en gris. Le ERDBD dans lequel Arg56 et Arg63 ont été remplacées par des alanines est représenté en noir. Le déplacement de H1 est de 1Å en moyenne. Des résultats similaires ont été observés après que les mêmes remplacements aient été effectués dans le GRDBD.

leurs interactions avec les groupements phosphates des HREs, fournissent un cadre structural pour les interactions spécifiques acides aminés/paires de bases.

En ce qui concerne le troisième acide aminé de la boîte P, dans le GRDBD Val29 fait un contact très important pour la sélectivité de la liaison à l'ADN en interagissant avec T-4. Ce contact établit une certaine distance de façon rigide entre H1 et le GRE. Dans le ERDBD, c'est une alanine qui se retrouve en position 29. Celle-ci faisant partie de la boîte P mais ne faisant aucun contact direct avec l'ADN, nous nous sommes interrogés à propos de son rôle potentiel dans le positionnement de H1 dans le sillon majeur du ERE. Pour ce faire, nous avons remplacé Ala29 par une glycine dans le ERDBD. Les expériences de minimisation d'énergie démontrent qu'effectivement, le remplacement de Ala29 par Gly entraîne un déplacement de 0,74Å de cette région de H1 vers le sillon majeur du ERE



(figure 3.7). Ces résultats démontrent que le troisième acide aminé de la boîte P pourrait potentiellement jouer un rôle dans le positionnement de H1 dans le sillon majeur de l'ADN.



**Figure 3.7 : Déplacement d'une partie de H1 du ERDBD après le remplacement de Ala29 par une glycine. Le ERDBD de type sauvage (wt) et le ERE sont représentés en gris. Le ERDBD dans lequel Ala29 a été remplacé par une glycine est représenté en noir. Le déplacement de la région centrale de H1 est d'environ 0,74Å.**

## **CHAPITRE 4 : DISCUSSION**

## **4.1 Modèle proposé expliquant le mécanisme de la sélectivité de la liaison à l'ADN par les récepteurs stéroïdiens**

### **4.1.1 Le modèle de Suzuki**

Ce modèle, applicable aux récepteurs stéroïdiens, était basé sur des règles générales de compatibilité chimique entre acides aminés et paires de bases ainsi que sur les contraintes stéréochimiques [27]. Ces contraintes ont été dérivées de la position de H1 dans le sillon majeur de l'ADN ainsi que de la longueur des acides aminés impliqués dans les interactions suite à l'étude des structures tridimensionnelles de divers récepteurs stéroïdiens. Ce modèle permettrait de prédire avec quelle séquence d'ADN un récepteur stéroïdien muté pourrait interagir. Par contre, ce modèle est basé sur les interactions un acide aminé/une base. Lorsque nous avons testé ce modèle, nos résultats démontrent que ce modèle ne représente pas la réalité et la complexité de l'interaction entre le ERDBD et le ERE (voir chapitre 2). Nos résultats mettent en lumière le rôle de Lys32, qui interagit avec la même paire de base que le premier acide aminé de la boîte P, Glu25, dans le cas du ERDBD. En effet, nous avons démontré que Lys32 et Glu25 jouent des rôles complémentaires dans la sélectivité de la liaison à l'ADN, ce qui aurait pour but d'augmenter la sélectivité de la liaison au ERE par le ERDBD. De plus, nos résultats nous poussent à croire que la mutation d'un seul résidu pourrait moduler le positionnement de H1 dans le sillon majeur de l'ADN et pourrait influencer la capacité du récepteur à reconnaître une séquence d'ADN ou non.

### **4.1.2 Modèle proposé tenant compte du positionnement de H1 dans le sillon majeur de l'ADN**

L'hypothèse de départ de la deuxième partie de ce projet de recherche était que la position de H1 dans le sillon majeur de l'ADN était un déterminant de la sélectivité de la liaison à l'ADN. Nous avons d'abord réussi à démontrer qu'il y avait une différence dans le positionnement de H1 dans le sillon majeur de l'ADN pour le GRDBD et le ERDBD (figure 3.1), deux récepteurs stéroïdiens de spécificité différente. Ces résultats sont en accord avec ce qui a déjà été observé [31]. Puisque les paires de bases centrales ont été

superposées, ceci implique que les molécules d'ADN sont dans des positions équivalentes et que c'est bel et bien H1 qui est positionné différemment dans le sillon majeur de l'ADN. Ces résultats peuvent aussi expliquer en partie pourquoi Lys32 du ERDBD peut contacter l'ADN et non dans le GRDBD (figure 1.8). En effet, selon la figure 3.1, H1 du ERDBD est beaucoup plus près de l'ADN que H1 du GRDBD.

Par la suite, nous voulions déterminer si cette différence dans la position de H1 entre les deux récepteurs avait vraiment un effet sur la sélectivité de la liaison à l'ADN. Pour ce faire, nous avons déplacé H1 du GRDBD dans la position occupée par H1 du ERDBD et avons vérifié si l'interaction entre le GRDBD et son élément de réponse était toujours possible. Nous avons aussi fait la même opération avec le ERDBD. Dans le cas du GRDBD, si H1 était dans la même position dans le sillon majeur que H1 du ERDBD, elle serait plus près de l'ADN tel que vu à la figure 3.1. Il était donc attendu que des collisions entre les atomes des acides aminés du GRDBD et des bases du GRE résultent de ce repositionnement. En effet, c'est ce que nous observons à la figure 3.2. Il est donc clair que dans ces conditions, il est impossible pour le GRDBD d'interagir avec le GRE en utilisant la même position de H1 dans le sillon majeur de l'ADN que le ERDBD. Dans le cas du ERDBD, le repositionnement de H1 dans la même position qu'occupe H1 du GRDBD aurait pour effet d'éloigner l'hélice  $\alpha$  de l'ADN comme montré à la figure 3.1. Il était donc attendu que certains ponts H entre le ERDBD et le ERE ne puissent plus se former. Le tableau III démontre effectivement que plusieurs ponts H ne peuvent plus se former entre les acides aminés du ERDBD et les bases du ERE. En conséquence, le petit nombre de ponts H formés entre le ERDBD et le ERE, lorsque H1 du ERDBD est dans la même position que H1 du GRDBD, ne peut promouvoir une forte interaction entre le ERDBD et le ERE.

Ces expériences simples de modélisation moléculaire nous ont donc permis d'établir que ce n'est pas seulement les séquences en acides aminés et en nucléotides qui déterminent la sélectivité de la liaison à l'ADN des récepteurs stéroïdiens, mais que c'est aussi la façon dont laquelle les récepteurs stéroïdiens interagissent avec leur HRE. En positionnant leur hélice de reconnaissance de l'ADN d'une manière spécifique dans le sillon majeur de leur HRE, les récepteurs stéroïdiens optimisent leur interaction avec l'ADN. Les récepteurs stéroïdiens doivent donc se soumettre à un critère de sélection

additionnel pour interagir correctement avec leur HRE et pour éviter d'interagir avec des HREs non-spécifiques.

L'analyse détaillée des complexes récepteur-ADN nous a amenés à redéfinir les rôles de certains acides aminés impliqués dans l'interaction avec l'ADN. Le positionnement de H1 dans le sillon majeur de l'ADN ainsi que les rôles joués par les différents acides aminés impliqués dans l'interaction protéine-ADN formeront la base du modèle que nous proposons pour expliquer la sélectivité de la liaison à l'ADN par les récepteurs stéroïdiens.

Nous voulions d'abord déterminer quels acides aminés étaient responsables du positionnement de H1 dans le sillon majeur de l'ADN, et ce, pour le GRDBD et le ERDBD. Deux groupes d'acides aminés étaient des candidats potentiels: les acides aminés de la boîte P, que nous savions déjà importants pour la sélectivité de la liaison à l'ADN [17 – 19], ou les acides aminés à l'extérieur de la boîte P mais interagissant l'ADN. Parmi ces derniers, il y a trois acides aminés basiques qui interagissent directement avec les bases des HREs : Lys28, Lys32 et Arg33 (figures 1.7 et 1.8). Lys32 interagit avec l'ADN seulement dans le complexe ERDBD-ERE et fait une interaction intramoléculaire avec Gln36 dans le GRDBD. Ceci peut s'expliquer en partie parce que H1 du GRDBD est positionné plus loin de l'ADN que H1 du ERDBD (figure 3.1). Nous avons réussi à démontrer, en remplaçant Lys28, Lys32 et Arg33 par des alanines dans les deux récepteurs et en faisant la minimisation d'énergie de ces complexes, que la position des H1 ne changeait pas de façon majeure. Seule, dans le cas du ERDBD, la partie C-terminale de H1 se rapprochait légèrement de l'ADN. Donc, de façon générale, ces trois acides aminés basiques ne sont pas impliqués dans le positionnement de H1 dans le sillon majeur de l'ADN sauf pour Lys32, dans le cas du ERDBD seulement. Nous suggérons que le rôle de Lys28 et Arg33 est d'augmenter l'affinité de l'interaction protéine-ADN. Ceci est en accord avec les études montrant que la mutation de Lys28 pour Ala dans le ARDBD [32] et la mutation de Arg33 pour Gly dans le GRDBD [33] réduisait considérablement (50%) ou abolissait complètement la liaison à l'ADN par ces récepteurs mutés. Pour ce qui est de Lys32, cet acide aminé ne semble pas jouer un rôle important dans la liaison au GRE par le GRDBD car, puisque la partie C-terminale de H1 du GRDBD est trop loin de l'ADN, il ne contacte pas l'ADN. Par contre, nos résultats nous permettent d'écarter la possibilité que Lys32 soit impliqué dans le positionnement de H1 du GRDBD dans le sillon majeur de l'ADN. Dans le cas du ERDBD, Lys32 semble jouer un rôle important pour la sélectivité de la liaison à

l'ADN pour ce récepteur. En effet, nos résultats montrent que la position de la partie C-terminale de H1 du ERDBD pourrait dépendre de la présence de Lys32. Cet acide aminé étant assez long, il peut interagir avec deux bases du ERE (figure 1.8) tout en maintenant la partie C-terminale de H1 à une certaine distance de l'ADN (figure 3.1). De plus, Lys32 contribue certainement à augmenter l'affinité de l'interaction ERDBD-ERE grâce aux deux ponts H formés. Finalement, nous avons démontré que Lys32 pouvait prévenir la liaison du ERDBD à des HREs non-spécifiques par conflit de charge ou encombrement stérique (figure 2.3). De plus, la présence de Lys32 suffit à maintenir la liaison à un ERE même lorsque le premier acide aminé de la boîte P, qui reconnaît les bases sur le brin opposé, est remplacé par une alanine (figure 2.5A). Pour conclure, nous croyons que la façon dont ces trois acides aminés basiques (Lys28, Lys32 et Arg33) interagissent avec l'ADN dépend du positionnement de H1 dans le sillon majeur de l'ADN. Ceci expliquerait pourquoi ces acides aminés conservés parmi les récepteurs stéroïdiens interagissent différemment avec l'ADN dans les complexes ERDBD et GRDBD (figure 1.7). Ceci implique aussi que le positionnement de H1 dans le sillon majeur de l'ADN dépend de l'identité des acides aminés de la boîte P.

#### **4.1.2.1 Rôle des acides aminés en deuxième position de la boîte P**

Effectivement, nos résultats démontrent que la présence de Gly ou Ser en position 26 déterminait le positionnement de la partie N-terminale de H1 grâce à leur entassement sur Arg56 et Arg63 (figure 3.5). Ces deux derniers acides aminés sont par ailleurs conservés à 100% parmi la famille des récepteurs nucléaires. Il a été démontré que le remplacement indépendant de ces deux arginines par une glycine abolissait complètement la liaison du GRDBD au GRE [33]. Il a aussi été rapporté que la mutation de Arg56 par une histidine dans le GRDBD causait la résistance au cortisol diagnostiquée chez un patient [34]. De plus, la mutation de Arg56 et Arg63 dans le ARDBD a été souvent rapportée comme une cause du syndrome d'insensibilité aux androgènes, un syndrome qui entraîne des problèmes de développement des organes sexuels mâles chez l'humain [35 – 39]. Ces résultats démontrent donc le rôle important que joue ces acides aminés. La figure 3.5 montre clairement le rôle joué par l'acide aminé en position 26. Puisque Arg56 et Arg63 sont presque parfaitement superposables entre les deux complexes et interagissent avec le

même groupe phosphate (figure 3.5), nous pouvons affirmer que ce ne sont pas les arginines qui causent la différence de positionnement de H1 dans le sillon majeur de l'ADN mais bien l'acide aminé en position 26 qui est entassé sur Arg56 et Arg63. Effectivement, la présence ou l'absence d'une chaîne latérale à cette position dicte la position de l'extrémité N-terminale de H1, comme démontré à la figure 3.5. Ceci explique pourquoi des acides aminés n'interagissant pas avec l'ADN sont tout de même définis comme étant des déterminants de la sélectivité de la liaison à l'ADN [17 – 19]. Le fait que Gly26 ou Ser26 s'entassent sur Arg56 et Arg63 pour déterminer la position de H1 dans le sillon majeur de l'ADN est en accord avec l'observation qu'il y a une contrainte pour de petits acides aminés (Gly, Ala, Ser, Thr et Cys) en position 26 [40]. En effet, nous pourrions imaginer que la présence d'un acide aminé en position 26 avec une grosse chaîne latérale (par exemple, Phe) serait incapable de positionner correctement H1 dans le sillon majeur de l'ADN et abolirait toute possibilité d'interaction avec l'ADN.

#### **4.1.2.2 Rôle des acides aminés en troisième position de la boîte P**

En ce qui concerne le troisième acide aminé de la boîte P (position 29), il s'agit d'une position importante pour la sélectivité de la liaison à l'ADN mais non équivalente dans les deux récepteurs [41]. Dans le GRDBD, la position 29 est occupée par une valine qui contacte T-4 dans le GRE (figure 1.8). L'importance de ce contact a été démontrée en remplaçant Val29 par une alanine dans le GRDBD. Ce remplacement a causé une baisse de l'affinité de la liaison au GRE de dix fois [28]. Ensuite, il a été démontré que le remplacement de T en position -4 (contacté par Val29) par U dans le GRE faisait aussi baisser l'affinité de liaison du GRDBD au GRE de dix fois [42]. Puisque Val29 est dans une conformation  $\alpha$ -hélicale, la conformation rotationnelle de la chaîne latérale de Val29 est restreinte de telle sorte qu'elle puisse contacter le groupement méthyl de T-4. En conséquence, il a été suggéré que ce contact devait dépendre de l'orientation de H1 dans le sillon majeur de l'ADN [20]. Nos résultats nous permettent d'ajouter que le contact entre Val29 et T-4 dépend aussi du positionnement de H1 dans le sillon majeur de l'ADN. Finalement, le rôle majeur de Val29 est de reconnaître la présence d'un T dans la séquence de l'ADN en un endroit précis. En plus de la reconnaissance du GRE, nous pouvons suggérer que le contact entre Val29 et T-4 maintient l'extrémité C-terminale de H1 à une

certain distance du GRE. Ceci empêcherait donc Lys32 d'interagir avec l'ADN. En conséquence, Lys32 préférerait interagir avec Gln36. Dans le ERDBD, la position 29 est occupée par une alanine ce qui élimine immédiatement la possibilité que le ERDBD puisse reconnaître le GRE, à la lumière de ce que nous venons de discuter. Ceci est en accord avec la suggestion que Ala29 n'augmentait pas l'affinité de la liaison à l'ADN mais prévenait plutôt la liaison à des HREs non correspondants [41]. Il a aussi été suggéré que, puisque la chaîne latérale de Ala est plus courte, H1 pourrait s'enfoncer plus profondément dans le sillon majeur de l'ADN [20] et effectivement, c'est ce que nous pouvons observer à partir de la figure 3.1. De plus, nos résultats montrent que la mutation de Ala29 par Gly dans le ERDBD permet à la région centrale de H1 de s'approcher encore plus de l'ADN (figure 3.7). Puisque H1 peut s'approcher plus près de l'ADN, Lys 32 du ERDBD peut donc interagir avec G+3 et T+4 (figure 1.8). En conclusion, nous suggérons que l'acide aminé en position 29, en plus d'être un élément clé dans la discrimination entre un GRE et un ERE, régule la position de l'extrémité C-terminale de H1 dans le sillon majeur de l'ADN et conséquemment, la capacité de Lys32 de contacter ou non l'ADN.

#### **4.1.2.3 Rôle des acides aminés en première position de la boîte P**

Finalement, la première position de la boîte P (position 25) est occupée par Gly dans le GRDBD et par Glu dans le ERDBD. Il a été démontré que ces acides aminés avaient des spécificités uniques [40, 43]. Dans le ERDBD, Glu25 interagit directement avec C-3 (figure 1.8), dirige un réseau de ponts H en concert avec des molécules d'eau et forme un pont salin avec Lys28 [21]. Ceci est le résultat des caractéristiques uniques de sa chaîne latérale qui permet donc à Glu25 d'interagir efficacement avec le C-3 spécifique au ERE et d'empêcher la liaison au GRE [40, 43]. Il a été récemment démontré que la mutation de Gly25 pour une arginine dans le ARDBD était la cause du syndrome de l'insensibilité aux androgènes diagnostiqué chez un patient. Des expériences de modélisation moléculaire ont montré que la présence de l'arginine en position 25 causait un encombrement stérique au niveau du carbone  $\beta$  de Arg25 et du groupement méthyle de T-4. Par conséquent, il a été suggéré qu'aucun acide aminé possédant un carbone  $\beta$  pouvait se retrouver en position 25, dans le contexte d'une hélice  $\alpha$  positionnée comme dans le complexe du GRDBD [44]. Ceci explique donc la présence d'une glycine en position 25 dans le GRDBD puisqu'il



s'agit du seul acide aminé à ne pas avoir de carbone  $\beta$ . De plus, il a été démontré que Gly25 était le seul acide aminé permettant au GRDBD de se lier à différents GREs, tout en empêchant la liaison au ERE [40]. En conclusion, l'acide aminé en première position de la boîte P est un important déterminant de la sélectivité de la liaison à l'ADN. Son double rôle consiste à lier efficacement les HREs correspondants et à empêcher la liaison à des HREs non correspondants. Nous croyons que la capacité de l'acide aminé en position 25 de jouer son rôle doit être fortement influencée par l'identité de l'acide aminé en position 26. Une mutation dans cette dernière position déplacerait H1 dans une position différente dans le sillon majeur de l'ADN et donc déplacerait aussi l'acide aminé en position 25. Celui-ci ne pourrait donc plus remplir son rôle adéquatement.

Pour conclure, nous croyons qu'il y a une forte synergie entre les trois acides aminés de la boîte P, puisque l'identité d'un seul acide aminé peut déterminer la capacité des autres acides aminés à remplir leur rôle. Ceci est particulièrement vrai dans le cas des acides aminés en deuxième et troisième position de la boîte P puisque nous avons démontré qu'ils étaient responsables du positionnement de H1 dans le sillon majeur de l'ADN. Cet aspect devrait donc être considéré lors de futures expériences de mutagenèse. En effet, une seule mutation pourrait influencer tout le mécanisme de la sélectivité de la liaison à l'ADN par la protéine et les résultats pourraient être plus difficiles à interpréter que prévu.

#### **4.2 Limites du modèle proposé**

Ce modèle expliquant le mécanisme de la sélectivité de la liaison à l'ADN par le GRDBD et le ERDBD peut s'appliquer aux récepteurs stéroïdiens (classe III du tableau I). Nous croyons que la dimérisation pourrait influencer la façon dont les récepteurs présentent leur H1 dans le sillon majeur de l'ADN. Puisque les récepteurs stéroïdiens forment des homodimères symétriques avec un arrangement en tête-à-tête, nous croyons que H1 se positionne de la même façon dans le sillon majeur de l'ADN pour les deux DBDs formant l'homodimère. Le modèle ne peut pas s'appliquer au ERR dans le cas où il se lie à un demi-site en tant que monomère. Puisque nous avons fait notre étude avec les structures de récepteurs sous forme de dimères, nous avons indirectement pris en compte les forces que la dimérisation pourraient exercer sur le positionnement de H1 dans le sillon majeur de l'ADN. La comparaison de la structure du ERR lié à l'ADN sous forme de monomère avec

les structures des complexes du ERDBD et du GRDBD devrait permettre de déterminer si notre modèle peut s'appliquer au ERR [45].

De plus, les récepteurs stéroïdiens lient des HREs qui ont la même séquence (excepté pour le ERE), dont les demi-sites sont orientés de la même manière et qui ont le même espacement entre les demi-sites. Pour pouvoir appliquer notre modèle expliquant la sélectivité de la liaison à l'ADN aux autres récepteurs nucléaires, il faudrait comparer les structures disponibles des récepteurs nucléaires d'autres types avec les structures du ERBDB et du GRDBD, pour vérifier si le positionnement de H1 peut être relié à la séquence d'ADN reconnue et à la boîte P du récepteur.

Pour l'instant, notre modèle ne peut s'appliquer que si les récepteurs lient les séquences consensus de HRE. La sélectivité de la liaison à l'ADN dépend fortement de la séquence du HRE reconnu par le récepteur. Il serait donc possible de faire la même analyse effectuée dans ce mémoire avec les séquences de HRE naturelles liées par les récepteurs stéroïdiens.

Il a été récemment découvert que le AR, qui est un récepteur stéroïdien, pouvait aussi lier des AREs dont les demi-sites étaient orientés en répétitions directes (DR) [6]. Nous ne sommes pas encore en mesure de déterminer si notre modèle peut s'appliquer dans ce cas-ci. Nous devons donc attendre la publication de nouveaux résultats comme la structure résolue par cristallographie aux rayons X de ce nouveau complexe ou encore la démonstration que d'autres récepteurs stéroïdiens peuvent lier des DRs. De plus, s'il s'avère que les récepteurs stéroïdiens peuvent lier des DRs, il serait intéressant de vérifier si les positions de H1 varient entre les récepteurs stéroïdiens et non-stéroïdiens lorsqu'ils sont liés à des DRs.

Dans le cas des récepteurs nucléaires formant des hétérodimères, sachant que la dimérisation peut influencer la façon dont le récepteur présente H1 dans le sillon majeur de l'ADN, une analyse approfondie des structures disponibles pourrait nous permettre de déterminer si notre modèle peut s'appliquer à ces récepteurs. Il a été récemment démontré pour le VDR qu'il existe de nouveaux déterminants de la sélectivité de la liaison à l'ADN à l'extérieur de la boîte P [46]. Ces nouvelles données viennent renforcer l'idée que les récepteurs formant des hétérodimères interagissent différemment avec l'ADN par rapport aux récepteurs stéroïdiens. Ces nouveaux déterminants de la sélectivité de la liaison à

l'ADN devront donc être pris en compte lors de l'analyse future de la structure de ces récepteurs.

### **4.3 Limites de la méthode**

Tout d'abord, l'analyse structurale se fait dans les limites imposées par la résolution des complexes structuraux analysés. Il s'agit de la marge d'erreur du modèle. Des complexes résolus à des résolutions plus petites pourraient permettre de rendre le modèle plus près de la réalité. Aussi, une structure résolue par cristallographie aux rayons X est fixée dans le cristal. Les complexes structuraux analysés ne tiennent donc pas compte de la dynamique interne de la molécule. De plus, le processus de minimisation d'énergie est artificiel puisque les complexes structuraux sont virtuellement dans le vide. Nous n'avons donc pas pris en compte les molécules d'eau entourant les complexes, les variations de la conformation de l'ADN (voir la section 3.1) et les ions pouvant se trouver en solution. Ces facteurs pourraient influencer le positionnement de H1 dans le sillon majeur de l'ADN, mais de façon minime. Des expériences de dynamique moléculaire pourraient nous permettre de considérer ces facteurs, quoique cette méthode prend beaucoup plus de temps et de ressources informatiques que la méthode que nous avons utilisée.

### **4.4 Conclusion et perspectives**

Pour conclure, nous avons réussi à démontrer que le mécanisme d'un positionnement différent de l'hélice de reconnaissance à l'ADN dans le sillon majeur de l'ADN consiste en un nouveau facteur de la sélectivité de la liaison à l'ADN. Ainsi, un modèle un acide aminé/une base expliquant la sélectivité de liaison à l'ADN en fonction des propriétés chimiques et stéréochimiques des intervenants a peu de chances de fonctionner, même dans le cadre réduit des récepteurs stéroïdiens interagissant avec leur élément de réponse palindromique. Nous avons alors proposé un modèle intégrant le rôle de chacun des acides aminés impliqués dans la sélectivité de la liaison à l'ADN et le positionnement de l'hélice de reconnaissance de l'ADN dans le sillon majeur de l'ADN. Ce modèle ne s'applique pour l'instant qu'aux récepteurs stéroïdiens, liés aux HREs de séquence consensus. Pour étendre ce modèle aux autres types de récepteurs nucléaires,

d'autres analyses structurales devront être effectuées avec la structure de ces récepteurs. De plus, pour intégrer d'autres facteurs ignorés jusqu'à présent comme la conformation de l'ADN, des techniques de dynamique moléculaire pourraient s'avérer efficaces, quoique plus coûteuses en terme de temps.

Finalement, ce mémoire montre que la sélectivité de la liaison à l'ADN par les récepteurs stéroïdiens est un mécanisme complexe dont les rouages ne sont pas encore complètement compris. Avec le nombre grandissant de complexes protéines-ADN dont la structure est publiée, le positionnement de l'hélice de reconnaissance de l'ADN comme facteur de sélectivité de la liaison à l'ADN pourra être analysé pour d'autres familles de protéines liant l'ADN.

## **BIBLIOGRAPHIE**

1. Renaud J.P. et Moras D. (2000) Structural studies on nuclear receptors. *Cell. Mol. Life Sci.* **57**, 1748 – 1769.
2. Schoneveld O.J., Gaemers I.C. et Lamers W.H. (2004) Mechanisms of glucocorticoid signalling. *Biochim. Biophys. Acta* **1680**, 114 – 128.
3. Gronemeyer H., Gustafsson J.-A. et Laudet V. (2004) Principles for modulation of the nuclear receptor superfamily. *Nat. Rev. Drug. Discov.* **3**, 950 – 964.
4. Sanchez R., Nguyen D., Rocha W., White J.H. et Mader S. (2002) Diversity in the mechanisms of gene regulation by estrogen receptors. *Bioessays* **24**, 244 – 254.
5. Aranda A. et Pascual A. (2001) Nuclear hormone receptors and gene expression. *Physiol. Rev.* **81**, 1269 – 1304.
6. Verrijdt G., Haelens A. et Claessens F. (2003) Selective DNA recognition by the androgen receptor as a mechanism for hormone-specific regulation of gene expression. *Mol. Genet. Metab.* **78**, 175 – 185.
7. Green S. et Chambon P. (1988) Nuclear receptors enhance our understanding of transcription regulation. *Trends Genet.* **4**, 309 – 314.
8. Hager G.L. Nagaich A.K., Johnson T.A., Walker D.D. et John S. (2004) Dynamics of nuclear receptor movement and transcription. *Biochim. Biophys. Acta* **1677**, 46 – 51.
9. Rastinejad F., Perlmann T., Evans R.M. et Sigler P.B. (1995) Structural determinants of nuclear receptor assembly on DNA direct repeats. *Nature* **375**, 203 – 211.
10. Klinge C.M. (2001) Estrogen receptor interaction with estrogen response elements. *Nucleic Acids Res.* **29**, 2905 – 2919.
11. Khorasanizadeh S. et Rastinejad F. (2001) Nuclear-receptor interactions on DNA-response elements. *Trends Biochem. Sci.* **26**, 384 – 390.
12. Beato M. et Klug J. (2000) Steroid hormone receptors: an update. *Hum. Reprod. Update* **6**, 225 – 236.
13. Luisi B.F., Xu W.X., Otwinowski Z., Freedman L.P., Yamamoto K.R. et Sigler P.B. (1991) Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* **352**, 497 – 505.
14. Mader S., Chambon P. et White J.H. (1993) Defining a minimal estrogen receptor DNA binding domain. *Nucleic Acids Res.* **21**, 1125 – 1132.

15. Gronemeyer H., Green S., Kumar V., Jeltsch J.-M. et Chambon P. (1988) Structure and function of the estrogen receptor and other members of the nuclear receptor family. In *Steroid receptors and disease: cancer, autoimmune, bone, and circulatory disorders*, P.J. Sheridan, K. Blum and M.C. Trachtenberg, édés. (New York: Marcel Dekker inc.), pp. 153 – 187.
16. Schwabe J.W.R. et Rhodes, D. (1991) Beyond zinc fingers: steroid hormone receptors have a novel structural motif for DNA recognition. *Trends Biochem. Sci.* **16**, 291 – 296.
17. Mader S., Kumar V., de Verneuil H. et Chambon P. (1989) Three amino acids of the oestrogen receptor are essential to its ability to distinguish an oestrogen from a glucocorticoid-responsive element. *Nature* **338**, 271 – 274.
18. Danielsen M., Hinck L. et Ringold G.M. (1989) Two amino acids within the knuckle of the first zinc finger specify DNA response element activation by the glucocorticoid receptor. *Cell* **57**, 1131 – 1138.
19. Umesono K. et Evans R.M. (1989) Determinants of target gene specificity for steroid/thyroid hormone receptors. *Cell* **57**, 1139 – 1146.
20. Arbuckle N.D. et Luisi B. (1995) A recipe for specificity. *Nat. Struct. Biol.* **2**, 341 – 346.
21. Schwabe J.W.R., Chapman L., Finch J.T. et Rhodes D. (1993) The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. *Cell* **75**, 567 – 578.
22. Garvie C.W. et Wolberger C. (2001) Recognition of specific DNA sequences. *Mol. Cell* **8**, 937 – 946.
23. Voet D. et Voet J.G. (1998) *Biochimie*, traduction de la 2ème édition américaine par Yves Gaudemer. (Paris, Bruxelles : De Boeck Université), pp. 175 – 176.
24. Glass C.K. (1994) Differential recognition of target genes by nuclear receptor monomers, dimers, and heterodimers. *Endocr. Rev.* **15**, 391 – 407.
25. Nelson C.C., Hendy S.C., Shukin R.J., Cheng H., Bruchovsky N., Koop B.F. et Rennie P.S. (1999) Determinants of DNA sequence specificity of the androgen, progesterone, and glucocorticoid receptors: evidence for differential steroid receptor response elements. *Mol. Endocrinol.* **13**, 2090 – 2107.
26. Schwabe J.W.R. (1997) The role of water in protein-DNA interactions. *Curr. Opin. Struct. Biol.* **7**, 126 – 134.
27. Suzuki M. et Yagi N. (1994) DNA recognition code of transcription factors in the helix-turn-helix, probe helix, hormone receptor, and zinc finger families. *Proc. Natl. Acad. Sci. USA* **91**, 12357 – 12361.

28. Alroy I. et Freedman L.P. (1992) DNA binding analysis of glucocorticoid receptor specificity mutants. *Nucleic Acids Res.* 20, 1045 – 1052.
29. Pabo C.O. et Nekludova L. (2000) Geometric analysis and comparison of protein-DNA interfaces: why there is no simple code for recognition? *J. Mol. Biol.* **301**, 597 – 624.
30. Pearlman D.A., Case D.A., Caldwell J.W., Ross W.R., Cheatham III T.E., DeBolt S., Ferguson D., Seibel G. et Kollman P. (1995) AMBER, a computer program for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to elucidate the structures and energies of molecules. *Comp. Phys. Commun.* **91**, 1 – 41.
31. Luisi B.F., Schwabe J.W.R. et Freedman L.P. (1994) The steroid/nuclear receptors: from three-dimensional structure to complex function. *Vitam. Horm.* **49**, 1 – 47.
32. Aarnisalo P., Santti H., Poukka H., Palvimo J.J. et Janne O.A. (1999) Transcription activating and repressing functions of the androgen receptor are differentially influenced by mutations in the deoxyribonucleic acid-binding domain. *Endocrinology* **140**, 3097 – 3105.
33. Hollenberg S.M. et Evans R.M. (1988) Multiple and cooperative trans-activation domains of the human glucocorticoid receptor. *Cell* **55**, 899 – 906.
34. Ruiz M., Lind U., Gafvels M., Eggertsen G., Carlstedt-Duke J., Nilsson L., Holtmann M., Stierna P., Wikstrom A.C. et Werner S. (2001) Characterization of two novel mutations in the glucocorticoid receptor gene in patients with primary cortisol resistance. *Clin. Endocrinol. (Oxf)* **55**, 363 – 371.
35. Schena M., Freedman L.P. et Yamamoto K.R. (1989) Mutations in the glucocorticoid receptor zinc finger region that distinguish interdigitated DNA binding and transcriptional enhancement activities. *Genes Dev.* **3**, 1590 – 1601.
36. Poujol N., Lobaccaro J.M., Chiche L., Lumbroso S. et Sultan C. (1997) Functional and structural analysis of R607Q and R608K androgen receptor substitutions associated with male breast cancer. *Mol. Cell. Endocrinol.* **130**, 43 – 51.
37. Pinsky L., Trifiro M., Kaufman M., Beitel L.K., Mhatre A., Kazemi-Esfarjani P., Sabbaghian N., Lumbroso R., Alvarado C., Vasiliou M., et Gottlieb B. (1992) Androgen resistance due to mutation of the androgen receptor. *Clin. Invest. Med.* **15**, 456 – 472.
38. Wooster R., Mangion J., Eeles R., Smith S., Dowsett M., Averill D., Barrett-Lee P., Easton D.F., Ponder B.A. et Stratton M.R. (1992) A germline mutation in the androgen receptor gene in two brothers with breast cancer and Reifenshtein syndrome. *Nat. Genet.* **2**, 132 – 134.



39. Gottlieb B., Lehvaslaiho H., Beitel L.K., Lumbroso R., Pinsky L. et Trifiro M. (1998) The Androgen Receptor Gene Mutations Database. *Nucleic Acids Res.* **26**, 234 – 238.  
La banque de données des mutations du gène du AR est accessible à:  
<http://www.mcgill.ca/androgendb/>
40. Zilliacus J., Carlstedt-Duke J., Gustafsson J.-A. et Wright A.P.H. (1994) Evolution of distinct DNA-binding specificities within the nuclear receptor family of transcription factors. *Proc. Natl. Acad. Sci. USA* **91**, 4175 – 4179.
41. Freedman L.P. (1992) Anatomy of the steroid receptor zinc finger region. *Endocr. Rev.* **13**, 129 – 145.
42. Truss M., Chalepakis G. et Beato M. (1990) Contacts between steroid hormone receptors and thymines in DNA: an interference method. *Proc. Natl. Acad. Sci. USA* **87**, 7180 – 7184.
43. Zilliacus J., Wright A.P.H., Carlstedt-Duke J. et Gustafsson J.-A. (1995) Structural determinants of DNA-binding specificity by steroid receptors. *Mol. Endocrinol.* **9**, 389 – 400.
44. Nguyen D., Steinberg S.V., Rouault E., Chagnon S., Gottlieb B., Pinsky L., Trifiro M. et Mader S. (2001) A G577R mutation in the human AR P box results in selective decreases in DNA binding and in partial androgen insensitivity syndrome. *Mol. Endocrinol.* **15**, 1790 – 1802.
45. Gearhart M.D., Holmbeck S.M., Evans R.M., Dyson H.J. et Wright P.E. (2003) Monomeric complex of human orphan estrogen related receptor-2 with DNA: a pseudo-dimer interface mediates extended half-site recognition. *J. Mol. Biol.* **327**, 819 – 832.
46. Hsieh J.C., Whitfield G.K., Jurutka P.W., Haussler C.A., Thatcher M.L., Thompson P.D., Dang H.T., Galligan M.A., Oza A.K. et Haussler M.R. (2003) Two basic amino acids C-terminal of the proximal box specify functional binding of the vitamin D receptor to its rat osteocalcin deoxyribonucleic acid-responsive element. *Endocrinology* **144**, 5065 – 5080.